



Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Laurea Magistrale in INGEGNERIA INFORMATICA

**Personalizzazione di head-related transfer
function basata su misure antropometriche**

Supervisor

STEFANO GHIDONI

Università di Padova

Co-supervisor

MICHELE GERONAZZO

Università di Aalborg

Master Candidate

LUCA BURIOLA

Abstract

Le tecnologie per la realtà virtuale hanno raggiunto un ottimo livello di maturità negli ultimi anni, e per creare un'immersività sempre maggiore occorrono miglioramenti anche nella componente audio legata alla spazializzazione del suono. Uno degli elementi chiave è la personalizzazione del contributo acustico dell'ascoltatore nella simulazione denominato head-related transfer function (HRTF). Tale contributo deve essere il più simile possibile a quello dell'utente-ascoltatore, dato che queste ultime non sono facilmente ricavabili. Occorre quindi un metodo per modellare le HRTF, e in questo lavoro di tesi viene adottato l'approccio a selezione di HRTF migliori. È stato esteso un approccio noto in letteratura per il quale la HRTF viene selezionata tramite contorni dell'orecchio e questo permetteva un miglioramento della localizzazione solo dei suoni in elevazione. Per migliorare ulteriormente la selezione quindi si è scelto di considerare anche le misure antropometriche della testa della persona per ottenere una modellazione più accurata dei ritardi temporali interaurali (informazioni contenute nella inter-aural time difference) più simile a quello reale, e tramite tecniche appropriate unire l'HRTF della selezione dell'orecchio a quest'ultima per ottenere una miglior localizzazione del suono. Per la validazione dei risultati è stato utilizzato un test di localizzazione, dove tramite un ambiente di realtà virtuale è stato chiesto, a una serie di soggetti, di individuare ripetutamente delle fonti sonore di diverse posizioni d'origine con i vari risultati dei metodi di selezione. Dai risultati ricavati da questo test è possibile concludere che il metodo implementato in questo lavoro migliora sensibilmente la localizzazione spaziale.

Indice

Abstract	v
Lista delle figure	xi
Lista delle tabelle	xiii
1 Introduzione	1
2 Concetti fondamentali del audio	7
2.1 Percezione spaziale del suono	8
2.1.1 Sistemi di coordinate	8
2.1.2 Percezione orizzontale del suono	9
2.1.3 Percezione dell'elevazione	10
2.1.4 Percezione della distanza	11
2.2 Head-related transfer function	12
2.2.1 Dominio delle HRTF	13
2.2.2 Modelli per l'HRTF	13
2.2.3 Misurazione di HRTF	14
2.3 ITD/ILD	17
2.3.1 Interaural time difference	17
2.3.2 Interaural level difference	18
2.4 HRTF database	19
2.4.1 CIPIC	19
2.4.2 Spatially Oriented Format for Acoustics	20
3 Background e stato dell'arte	23
3.1 Obbiettivi della tesi	23
3.2 Selezione di HRTF basata sui contorni	24
3.2.1 Mismatch in frequenza	24
3.2.2 Mismatch antropometrico	27
3.2.3 Classificazione dei risultati	28
3.3 Modelli per ITD	29
3.3.1 Modello sferico	29
3.3.2 Modello sferico con la traslazione delle orecchie	30
3.3.3 Modello ellissoidale	32

3.4	Tecniche di estrazione e modifica di ITD	35
3.4.1	Estrazione	35
4	Implementazione del software di personalizzazione	39
4.1	Interfaccia grafica	39
4.1.1	Opzioni d’inserimento dati	40
4.1.2	Rilevamento dei dati per l’elaborazione	41
4.2	Metriche di confronto	42
4.3	Supporto SOFA	44
4.3.1	GUI per la creazione di SOFA	44
4.4	Uso del Plugin	45
5	Test di localizzazione	47
5.1	Descrizione del test di localizzazione	47
5.1.1	Hardware & software	47
5.1.2	Ambiente	49
5.2	Motivazione dei confronti	50
5.2.1	Metriche scelte	50
5.2.2	Confronto long e short	50
5.2.3	A_{opt} e Cross-correlation	51
5.2.4	Metodi combinati	51
5.3	Protocollo sperimentale	52
6	Risultati	55
6.1	L’errore di localizzazione	56
6.2	Valutazione sull’errore medio	56
6.2.1	Confronto Short-Long	56
6.2.2	Confronti tra selezioni	58
6.2.3	Errori di Front-Back	61
6.2.4	Eliminazione delle anomalie	61
6.3	Risultati Nasa TLX	62
7	Conclusioni	67
	Bibliografia	69

Elenco delle figure

1.1	Sistema audio surround di tipo 5.1 [1]	2
2.1	Confronto tra il sistema verticale(a) e interneurale(b) [2]	9
2.2	Esempio di forma d'onda arrivata alle due orecchie. [3]	10
2.3	Cono di confusione [4]	11
2.4	Plot di una HRTF e HRIR per un determinato azimuth ed elevazione	12
2.5	Configurazione per la misura di HRTF all'Acoustics Research Institute, Austrian Academy of Sciences, Vienna. [5]	15
2.6	Punti di misurazione del database CIPIC. [6]	20
2.7	Misure di testa e torso. [6]	21
2.8	Misure dell'orecchio. [6]	21
2.9	Esempio di file SOFA aperto utilizzando HDF5View. Tra i vari para- metri troviamo Data.IR che racchiude la risposta impulsiva e Source- Position che descrive le posizioni spaziali dei relativi punti.	22
3.1	Rappresentazione dei cinque contorni principali. [7]	26
3.2	Rappresentazione delle orecchie spostate. [8]	31
3.3	Performance del modello sferico con orecchie spostate . Le linee rosse rappresentano l'udibilita' dell'errore, quelle grige l'errore dei soggetti, e quella nera la media degli errori. [8]	33
3.4	Dimensioni della testa e posizione delle orecchie. [9]	34
3.5	Dimensioni della testa e posizione delle orecchie. [8]	35
4.1	Schermata della GUI prima delle modifiche. [10]	40
4.2	Schermata della GUI dopo delle modifiche.	40
4.3	Schermata del menu' di generazione dei file Sofa. [10]	45
4.4	Schermata del menu' di generazione dei file Sofa. [10]	46
5.1	Schermata di Unity rappresentante lo scenario del test di localizzazione	48
5.2	Schermata di Steam audio manager all'interno di Unity [11]	49
5.3	Soggetto all'interno della camera silente pronto per effettuare il test. .	53
6.1	Tabella dei soggetti partecipanti al test.	55
6.2	Istogramma dei dati di errore medio assoluto per l'azimuth.	60
6.3	Istogramma dei dati di errore medio assoluto per l'elevazione.	60
6.4	NasaTLX Soggetto 1	63

6.5	NasaTLX Soggetto 2	63
6.6	NasaTLX Soggetto 3	63
6.7	NasaTLX Soggetto 4	63
6.8	NasaTLX Soggetto 5	63
6.9	NasaTLX Soggetto 6	63
6.10	NasaTLX Soggetto 7	64
6.11	NasaTLX Soggetto 8	64
6.12	NasaTLX Soggetto 9	64
6.13	NasaTLX Soggetto 10	64
6.14	NasaTLX Soggetto 11	64
6.15	NasaTLX Soggetto 12	64
6.16	NasaTLX Soggetto 13	64
6.17	NasaTLX Soggetto 14	64
6.18	Grafico per la media dei risultati dei Nasa TLX	65

Elenco delle tabelle

3.1	Confronto attraverso media assoluta e media quadratica tra il modello sferico con e senza traslazione e quello ellittico per i parametri diretti, predetti e ottimi. [8]	32
3.2	Confronto tra i vari metodi di estrazione di ITD [12].	36
6.1	Errore medio assoluto di azimuth e di elevazione tra kemar e cipic sia per i test long che short.	57
6.2	Deviazione standard di azimuth e di elevazione tra kemar e cipic sia per i test long che short.	57
6.3	Risultati delle 3 ripetizioni per Kemar e la selezione dei contorni. . . .	58
6.4	Errore medio assoluto nella rilevazione del azimuth per le varie selezioni.	59
6.5	Errore medio assoluto nella rilevazione dell'elevazione per le varie selezioni.	59
6.6	Deviazione standard nella rilevazione del azimuth per le varie selezioni.	59
6.7	Deviazione standard nella rilevazione dell'elevazione per le varie selezioni.	60
6.8	Deviazione standard nella rilevazione dell'elevazione per le varie selezioni.	60
6.9	Numero di front-back per ogni condizione e per ogni soggetto.	61
6.10	Errore medio assoluto nella rilevazione del azimuth per le varie selezioni eliminando i soggetti 4,10 e 12.	61
6.11	Errore medio assoluto nella rilevazione dell'elevazione per le varie selezioni eliminando i soggetti 4,10 e 12.	62
6.12	Deviazione standard nella rilevazione del azimuth per le varie selezioni eliminando i soggetti 4,10 e 12.	62
6.13	Deviazione standard nella rilevazione dell'elevazione per le varie selezioni eliminando i soggetti 4,10 e 12.	62
6.14	Tabella per la media e varianza dei risultati dei Nasa TLX	65

1

Introduzione

L'intrattenimento multimediale rappresenta, dalla sua creazione in poi, uno dei maggiori campi per l'interesse e lo svago delle persone. Questo si traduce in costanti investimenti e ricerche in questi ambienti volti ad aumentare il più possibile i profitti [13].

Oltre agli ovvi sviluppi in ambito grafico negli ultimi decenni è stato possibile osservare una crescente attrattiva per i sistemi di riproduzione multicanale. Le sale cinematografiche incorporano sistemi di diffusione più perforanti, per consentire all'ascoltatore un'esperienza più coinvolgente. Inoltre, i cosiddetti sistemi home cinema, insieme agli standard di audio surround come 5.1 o 10.2, permettono la spazializzazione del suono a 360 gradi [1].

In questo caso la spazializzazione viene creata da strumenti di post-produzione o tramite effetti. In altre parole, la determinazione della traiettoria spaziale di un suono non fa sì che il suono venga spazializzato istantaneamente, così come, ad esempio, la scrittura di una partitura non fa sì che la musica venga ascoltata immediatamente. Per quanto concerne l'ambiente videoludico, l'audio è stato una parte cruciale dell'esperienza di gioco fin dall'avvento dei primi giochi arcade [14]. Nel corso del tempo, lo stato dell'audio digitale è costantemente migliorato, dai semplici generatori d'onda alla sintesi FM, evolvendosi da sample mono a 8 bit e sample stereo a 16 bit, fino agli attuali sistemi audio surround 5.1 delle moderne console di gioco [14]. Dallo sviluppo del surround 5.1, poco è cambiato, la tecnologia fondamentale è la stessa e l'ambiente di

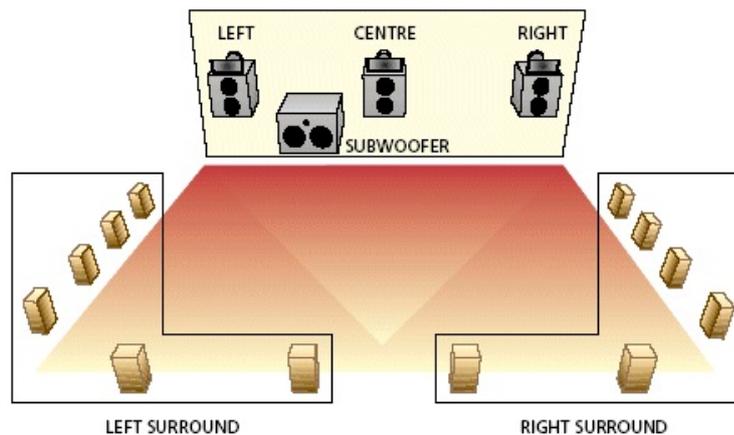


Figura 1.1: Sistema audio surround di tipo 5.1 [1]

gioco è ancora principalmente il soggiorno con un grande televisore e altoparlanti. Per quanto riguarda la realtà virtuale, tuttavia, le cose sono diverse. Invece di un grande ambiente con altoparlanti, la realtà virtuale porta l'esperienza vicino all'utente attraverso un head-mounted display (HMD) per la parte visiva e delle cuffie per la parte sonora. La capacità di tracciare l'orientamento e la posizione della testa dell'utente consente di potenziare in modo significativo la tecnologia audio. Finora, l'attenzione è sempre stata posta tipicamente sugli aspetti visivi della realtà virtuale, ma anche l'audio è importante per fornire il massimo senso di immersione possibile. La necessità dell'audio negli ambienti virtuali, ha spinto molti colossi dell'informatica verso la creazione di motori audio sempre più potenti e sempre più personalizzabili. Tra questi troviamo ad esempio:

- Google LLC¹: il suo motore si chiama Resonance Audio, un SDK audio spaziale multiplatforma, che offre alta fedeltà del audio spazializzato. Questa potente tecnologia è fondamentale per esperienze realistiche per AR, VR, giochi e video. È possibile utilizzare Resonance Audio in una varietà di strumenti di gioco, audio e di sviluppo di applicazioni e distribuirlo su più piattaforme come Unity, Unreal, FMOD e Wwise.
- Oculus VR²: nella sua SDK per lo sviluppo in ambito realtà virtuale, mette a disposizione audio SDK, che permette la realizzazione di audio spazializzato per

¹<https://resonance-audio.github.io/resonance-audio/>

²<https://developer.oculus.com/audio/>

varie piattaforme come: Unity, Unreal Engine 4, Audiokinetic Wwise, FMOD Studio, Avid Pro-Tools (AAX), Various Windows DAWs (VST).

- Valve Corporation³: Offre Steam Audio, una SDK pensata appositamente per la spazializzazione dell'audio. Steam Audio permette una soluzione audio completa che integra la simulazione acustica dell'ambiente e dell'ascoltatore. La propagazione sonora basata sulla fisica ricrea costantemente il modo in cui il suono interagisce con l'ambiente virtuale.

Questi software permettono di gestire i seguenti aspetti del suono:

- spazializzazione: trasformare le sorgenti sonore monofoniche per riprodurle come se provenissero da una specifica direzione.
- modellazione della propagazione del suono: Fornire una simulazione in tempo reale di riverbero e occlusione basata sulla geometria del ambiente virtuale circostante.
- realtà virtuale: algoritmi per il tracciamento rotazionale e posizionale, audio basato sulla fisica a bassa latenza e supporto di un'ampia gamma di hardware per creare ambienti immersivi di realtà virtuale (VR).
- integrazione: l'integrazione con i principali motori di gioco e middleware audio standard del settore, e il supporto di diverse piattaforme per lo sviluppo in VR.

Determinato dunque l'importanza della parte sonora in questi ambienti, andiamo a vedere com'è possibile la spazializzazione di questi ultimi. Questa ricopre una grande parte dell'audio VR: è la capacità di riprodurre un suono come se fosse posizionato in un punto specifico dello spazio tridimensionale. La spazializzazione è un aspetto chiave della presenza, cioè la sensazione di essere presenti nel ambiente, perché fornisce potenti spunti che suggeriscono che all'utente di trovarsi in un vero e proprio ambiente 3D, il che contribuisce fortemente al senso di immersione.

Ci sono due componenti chiave per la localizzazione di una sorgente sonora nello spazio [14]: la direzione e la distanza.

Le persone usano diversi indicatori per dedurre la distanza da una sorgente sonora. Questi sono:

³<https://valvesoftware.github.io/steam-audio/>

- intensità: è una delle caratteristiche di maggior importanza, e si modella tramite una semplice attenuazione del segnale basata sulla distanza tra la sorgente e l'ascoltatore in base alla dissipazione energetica dovuta alla propagazione del suono nell'aria.
- initial time delay: descrive l'intervallo tra il suono diretto e la sua prima riflessione. Più lungo è questo intervallo, più ci si avvicina alla sorgente sonora.
- rapporto tra suono diretto e riverbero: in un ambiente riverberante è presente una lunga e diffusa coda sonora composta da tutti gli echi tardivi che interagiscono tra loro. Più si sente il suono diretto rispetto al suono riverberato, più si è vicini alla fonte.
- motion parallax: rappresenta il movimento apparente di una sorgente sonora attraverso lo spazio. Questo risulta maggiore per fonti vicine all'ascoltatore e minore per quelle più lontane.
- Attenuazione ad alta frequenza: le alte frequenze si attenuano più velocemente delle basse frequenze, quindi su lunghe distanze possiamo dedurre la distanza in base a quanto sono attenuate.

Per determinare la direzione l'ascoltatore utilizza principalmente:

- interaural time difference (ITD): il ritardo tra l'arrivo del suono ad entrambe le orecchie.
- interaural level difference (ILD): la differenza nel livello di volume del suono a entrambe le orecchie.
- head-related transfer function (HRTF): Un filtro per le trasformazioni spettrali dovute alla direzione, codificato come funzione di trasferimento, basata sulle misure reali del corpo.

In questa tesi ci baseremo sullo studio e sulla personalizzazione degli indicatori di direzionalità.

L'approccio più comune in un'applicazione di realtà virtuale è l'uso di un set generico di HRTF misurata con una *dummy head* che rappresenta una media di un'ampia fetta della popolazione umana [15]. Tuttavia, l'uso di tali set di dati e, in generale, l'uso di HRTF non individuali, introduce un aumento degli errori di localizzazione di fonti audio, confusione front-back e mancanza di esternalizzazione. Per porre rimedi a questi

errori si può ricorrere all'uso di HRTF "costruite" sulla base delle dimensioni antropometriche dell'utente. Ovviamente la scelta migliore sarebbe quella di calcolare l'HRTF individuale di ogni utente. Questo purtroppo risulta spesso molto complicato, se non addirittura impossibile. Un approccio alternativo utilizza funzioni di trasferimento estratte da database che assomiglino il più possibile a quelle personali dell'utente.

Lo scopo di questa tesi è infatti quello di scegliere da un database di HRTF già disponibili quella più simile alla HRTF dell'utente secondo specifiche misure ricavate dalla fisionomia dell'orecchio, e andare a modificarla sostituendo il suo ITD con quello di un'altra HRTF selezionata dallo stesso database tramite le misure antropometriche della testa, per creare un'unica HRTF partendo dalle due selezioni. Il lavoro di selezione basato sulla forma dell'orecchio e dei suoi contorni è già stato svolto in precedenza [16] [10], e mette a disposizione un software per la scelta e la selezione dell'HRTF. Si userà quindi quest'ultimo come base di partenza, per poi implementare le ulteriori modalità di selezione descritte.

La tesi è strutturata come segue:

Nel primo capitolo si introducono le nozioni generali per comprendere i concetti trattati su questa tesi, nel secondo capitolo vengono invece analizzate le tecniche specifiche per la realizzazione del lavoro. Nel terzo capitolo si parla dello svolgimento pratico, nel quarto si va a vedere quali sono i confronti per provare la veridicità di quanto scritto. Infine negli ultimi due capitoli vengono presentati i risultati ottenuti e le conclusioni finali di questa tesi.

2

Concetti fondamentali del audio

Descriviamo in questo capitolo i concetti fondamentali che serviranno per comprendere al meglio il lavoro di questa tesi. Partiamo quindi dal concetto fondamentale di suono.

Il suono è un disturbo meccanico o vibrazione, che si propaga attraverso un materiale elastico. In pratica il suono è un'onda, detta onda sonora che oscilla intorno ad una posizione di equilibrio lungo la sua direzione di propagazione [17].

Le caratteristiche principali del suono sono:

- altezza: può essere definita come quell'attributo uditivo attraverso il quale i suoni possono essere ordinati su una scala musicale che va dall'alto al basso. Come il volume e il timbro, è un attributo soggettivo, che non può essere espresso in unità fisiche.
- intensità: è definita come quella sensazione uditiva che più corrisponde alla misura fisica dell'intensità sonora. Questa dipende fortemente sia dall'intensità fisica che dal suo contenuto in frequenza.
- timbro: è legato alla struttura armonica del suono. rappresenta quell'attributo della sensazione uditiva che consente all'ascoltatore di identificare la fonte sonora, rendendola distinguibile da ogni altra.

2.1 Percezione spaziale del suono

2.1.1 Sistemi di coordinate

Data la definizione di suono passiamo ora ad un altro concetto fondamentale per localizzare un suono nello spazio. Un sistema di coordinate è un sistema che utilizza uno o più numeri, o coordinate, per determinare in modo univoco la posizione dei punti su uno spazio [3].

I due sistemi che vengono maggiormente usati sono:

- cartesiano: Vengono scelte due linee nel piano, e le coordinate vengono calcolate dalla distanza del punto rispetto agli assi
- sferico: Un sistema di coordinate sferiche è un sistema di coordinate per lo spazio tridimensionale dove la posizione di un punto è specificata da tre numeri: la distanza di quel punto dall'origine, detto raggio, il suo angolo polare che misura la sua "altezza", detto elevazione, e l'angolo che misura la distanza tra la retta di riferimento dell'asse e quella che collega il punto e l'origine. Può essere vista come la versione tridimensionale del sistema di coordinate polari.

Per quanto riguarda il sistema sferico, due tipi sono maggiormente utilizzati, quello verticale e quello interaurale [2].

- vertical-polar: In questo sistema rappresentato in Fig.2.1(a) l'azimuth è misurato come l'angolo che va dal piano yz ad un piano verticale contenente il punto sorgente e l'asse z, l'elevazione è poi l'angolo che parte dal piano xy. Con questa scelta, le superfici di azimuth costanti sono piani che attraversano l'asse z mentre quelle di elevazioni costanti sono coni concentrici intorno all'asse z.
- interaural-polar: In questo caso rappresentato in Fig.2.1(b) l'elevazione viene misurata come angolo dal piano xy ad un piano contenente il punto sorgente e l'asse x, l'azimut viene poi misurato come angolo dal piano yz. Con questa scelta, le superfici di elevazione costante sono piani che attraversano l'asse x mentre le superfici di azimuth costante sono coni concentrici con l'asse x.

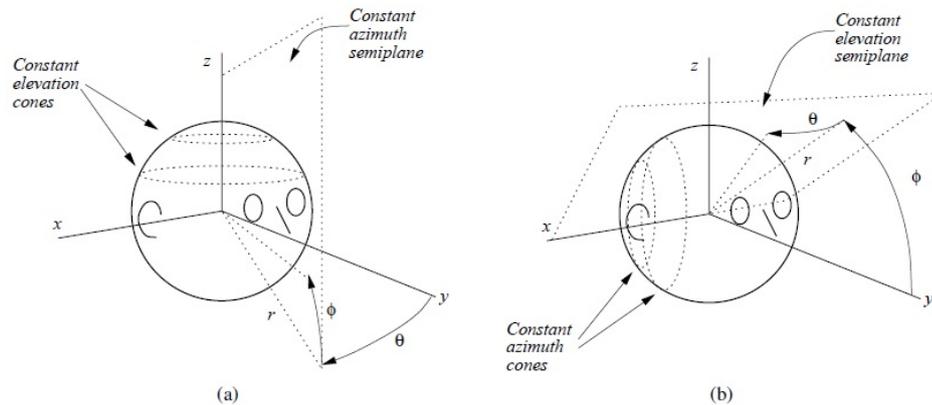


Figura 2.1: Confronto tra il sistema verticale(a) e interneurale(b) [2]

2.1.2 Percezione orizzontale del suono

Il posizionamento orizzontale delle orecchie massimizza le differenze delle fonti sonore che si verificano intorno all'ascoltatore, piuttosto che dal basso o dall'alto, consentendo l'ascolto di suoni che occorrono a livello del terreno e di quelli esterni al campo visivo. L'ITD e l'ILD sono considerati i parametri chiave per la percezione dell'azimuth, in quella che viene chiamata Teoria Duplex della localizzazione [4].

Consideriamo quindi un'onda che raggiunge l'orecchio destro e sinistro. Alle basse frequenze la ITD è identificabile dallo spostamento della forma d'onda di una frazione di ciclo, che è facilmente rilevabile come si può vedere in Fig2.2(a). Qualitativamente si può dire che se la metà della lunghezza d'onda è più grande delle dimensioni della testa, allora è possibile per l'uditivo rilevare la fase di queste forme d'onda in modo inequivocabile e quindi l'effetto dell'ITD può essere avvertito. Nel caso di alte frequenze invece, c'è ambiguità nel ITD, poiché ci possono essere diversi cicli di spostamento come accade in Fig2.2(b).

Possiamo considerare il punto critico come il punto in cui la metà della lunghezza d'onda risulta più corta della dimensione della testa: per lunghezze d'onda più corte, non si può più determinare quale sia il fronte d'onda principale. Il punto critico in frequenza è di solito assunto come un valore intorno a 1,5 kHz.

Se ora guardiamo l'ILD, la situazione è invertita. Alle basse frequenze la funzione di trasferimento della testa è essenzialmente piatta e quindi ILD risulta poco influente.

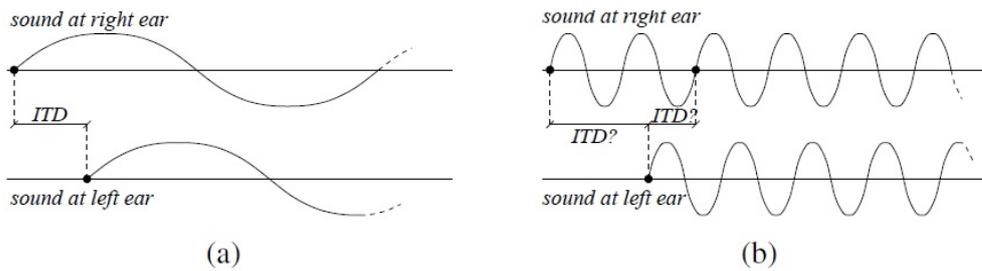


Figura 2.2: Esempio di forma d'onda arrivata alle due orecchie. [3]

D'altra parte, alle alte frequenze l'ILD è più marcato e può diventare molto grande. Per questo motivo la Teoria del Duplex afferma che l'ILD e l'ITD sono complementari, e quindi insieme coprono sia le basse che le alte frequenze.

Front-Back confusion

Le informazioni fornite da ITD e ILD possono essere ambigue. Prendendo ad esempio un suono che arriva all'ascoltatore da un punto θ , e un secondo suono proveniente esattamente da dietro in posizione $\pi - \theta$ i valori di ITD e ILD saranno gli stessi [4]. Questo genera quindi una confusione nel catalogare i suoni provenienti da di fronte come se fossero provenienti dalle spalle o viceversa.

2.1.3 Percezione dell'elevazione

Possiamo immaginare che esista una circonferenza formata da punti equidistanti dall'orecchio dove ognuno di questi punti presenta valori equivalenti sia di ITD che di ILD, come rappresentato in Fig.2.3. Queste circonferenze sono spesso indicate come coni di confusione, ed estendono il concetto di Front-Back confusion che visto in precedenza. I padiglioni auricolari possono disambiguare questa confusione, e per questo vengono considerati particolare importante per la localizzazione verticale [4]. Per quanto riguarda le caratteristiche spettrali, l'angolo di elevazione causa lo spostamento dei picchi e dei notches in funzione dell'orientamento sia della fonte sonora che dell'ascoltatore.

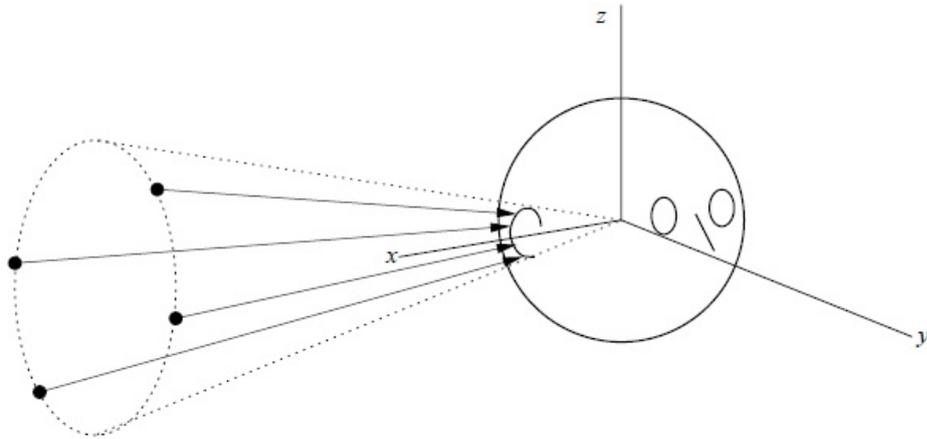


Figura 2.3: Cono di confusione [4]

2.1.4 Percezione della distanza

L'intensità è la caratteristica primaria per calcolare la distanza utilizzato dagli ascoltatori, che imparano attraverso l'esperienza la distanza dalle sorgenti sonore e i loro spostamenti attraverso aumenti o riduzioni d'intensità. La riduzione dell'intensità sonora con l'aumentare della distanza segue un andamento quadratico inverso: l'intensità di una sorgente sonora diminuirà approssimativamente di 6 dB ogni volta che raddoppia la distanza [4]. Tuttavia questa legge esprime il rapporto tra l'intensità di una sorgente sonora e un livello di riferimento, mentre la grandezza di intensità percepita viene chiamata loudness. Gli incrementi di loudness possono funzionare efficacemente come indicazioni di distanza solo in assenza di altre informazioni sonore, in particolare il riverbero. Quando il riverbero è presente il volume complessivo all'orecchio dell'ascoltatore non cambia molto per sorgenti molto vicine o molto lontane: la dipendente dalla distanza si applica solo al suono diretto mentre l'energia riflessa rimane approssimativamente costante. In questo caso quindi, il rapporto tra suono diretto e suono riflesso diventa un fattore più importante della per la localizzazione rispetto alla variazione di intensità del suono.

2.2 Head-related transfer function

La HRTF, rappresentata in Fig. 2.4, descrive come una data onda sonora in ingresso è filtrata dagli effetti di defrazione e riflessione creati dalle caratteristiche fisiche della testa, del padiglione e del busto, prima che il suono raggiunga i timpani dell'ascoltatore. Definiamo formalmente l'HRTF ad un orecchio come il rapporto di frequenza tra il livello di pressione sonora (SPL) $\Phi^{(l),(r)}(\theta, \phi, \omega)$ al timpano destro o sinistro e l'SPL al che si riscontrerebbe al centro della testa $\Phi_f(\omega)$ se l'ascoltatore non fosse presente [9]:

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

L'HRTF è una funzione in frequenza mentre utilizzando la trasformata di Fourier inversa, si ottiene la Head-related impulse response (HRIR), visualizzabile in Fig. 2.4, che è la funzione nel tempo corrispondente. Queste due funzioni insieme incorporano le informazioni riguardante l'ITD e l'ILD, ed è quindi possibile estrarle tramite tecniche appropriate. Le HRTF nella maggior parte dei casi presentano valori spettrali molto simili nei vari individui, ma date le varie differenze antropometriche tra le persone, tra una e l'altra troviamo sottili differenze. Queste differenze individuali hanno dimostrato di giocare un ruolo importante per una localizzazione precisa [15].

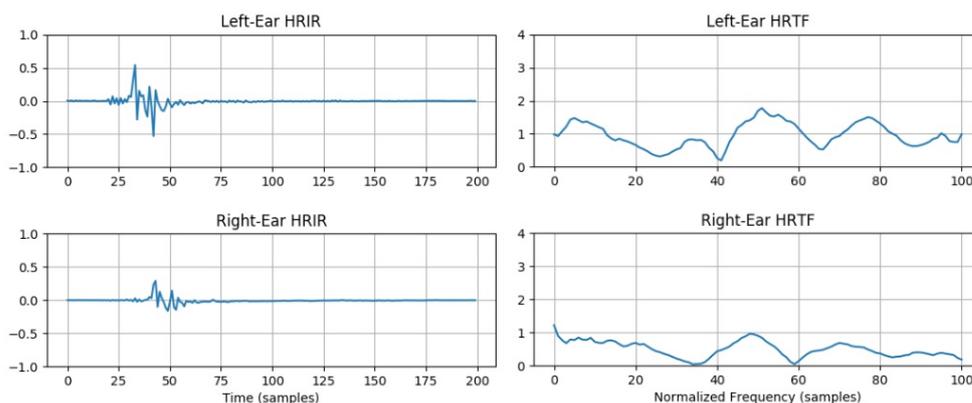


Figura 2.4: Plot di una HRTF e HRIR per un determinato azimuth ed elevazione

2.2.1 Dominio delle HRTF

La HRTF può essere rappresentata nei domini temporali, di frequenza e spaziale.

Dominio delle frequenze

I valori che assume l'HRTF nel dominio della frequenza [18] sono rappresentati dalla magnitudine e dalle risposte di fase in diverse posizioni in funzione della frequenza. La visualizzazione HRTF nel dominio della frequenza mostra alcune proprietà dei dati, come picchi, notches o altre particolari forme spettrali.

Dominio del tempo

Nel dominio del tempo l'HRTF viene rappresentata attraverso la sua corrispettiva HRIR [18]. Questa viene rappresentata attraverso la variazione della sua ampiezza nel tempo nelle varie posizioni di azimuth ed elevazione.

Dominio dello spazio

La visualizzazione spaziale dei dati HRTF può essere presentata in diversi modi [18]. Un esempio di ciò è il plot di solo ITD e ILD in funzione dell'azimuth e dell'elevazione. Più in generale, si traccia l'andamento dei valori nel dominio di tempo o frequenza, mantenendo fisso uno dei due valori di azimuth o elevazione, facendo variare l'altro.

2.2.2 Modelli per l'HRTF

I modelli computazionali permettono la creazione di HRTF sintetiche. Questi modelli hanno diversi gradi di semplificazione, si varia da modelli con semplici assunzioni geometriche fino a descrizioni più accurate, in grado di riprodurre i picchi e i notches della funzione.

Negli ultimi due decenni sono state proposte diverse tecniche per la progettazione di HRTF e queste possono essere classificati in quattro categorie [19]:

- modelli polo/zero: si utilizzano varie tecniche per adattare i modelli parametrici alle misurazioni effettuate per creare HRTF approssimate con filtri di basso ordine.
- espansioni in serie: si utilizza una somma ponderata di funzioni più semplici applicate ad una serie di HRIR o HRTF.

- modelli strutturali: vengono calcolati singolarmente i dati antropometrici dell'ascoltatore. La linearità di questi contributi permette di ricostruire l'HRTF globale a partire da una combinazione di tutti questi effetti.
- modello strutturale misto [19]: segue l'approccio di modellazione strutturale, generalizzandolo in modo da consentire l'inserimento di misurazione acustiche, simulazioni numeriche, o uno degli approcci di modellazione in uno o più dei suoi componenti spettrali.

2.2.3 Misurazione di HRTF

Un approccio comune per misurare l'HRTF è quello di inserire un microfono vicino o all'interno del condotto uditivo di un utente, ed emettere un segnale dello spettro noto attraverso un altoparlante posto in una determinata posizione di azimuth, elevazione e distanza dalla testa del soggetto. Per tale misura, il soggetto è solitamente posto al centro del dispositivo di misura, sul quale possono essere posizionati altoparlanti fissi o mobili nella posizione specificata.

La misura viene solitamente effettuata in una camera anecoica. Un esempio del settaggio è riportato in Fig.2.5. Poiché l'HRTF è un sistema LTI, per la misurazione vengono utilizzati i metodi standard di identificazione dei sistemi LTI. La risposta all'impulso misurata viene poi troncata in corrispondenza di un certo intervallo di tempo in modo che tutti i riverberi successivi vengono scartati. Per compensare le distorsioni lineari introdotte dagli altoparlanti e dai microfoni, viene effettuata una misura di riferimento senza soggetto con il microfono posto nell'origine, e questa, viene rimossa dalle misure grezze. Ci sono ancora molti altri fattori che possono degradare la fedeltà di la misura HRTF, ad esempio rumori acustici ed elettrici, movimenti del soggetto e imprecisioni degli apparecchi. Tutti questi fattori devono essere presi in considerazione nella progettazione del sistema. La necessità di utilizzare grandi apparecchiature e la difficoltà di calcolo appena presentata, rende molto ostico l'acquisizione di HRTF personali per ogni singolo utente.

Selezione di HRTF

Un'HRTF personalizzata viene di solito scelta da un database di HRTF misurate su diversi soggetti attraverso le tecniche precedentemente viste. In questo caso quindi si



Figura 2.5: Configurazione per la misura di HRTF all'Acoustics Research Institute, Austrian Academy of Sciences, Vienna. [5]

utilizzeranno procedure diverse per la selezione della funzione più simile tra l'utente e un soggetto del database. Il vantaggio principale di questa tecnica risulta abbastanza ovvio, infatti, non c'è la necessità di dover acquisire l'HRTF individuale della persona, che come visto in precedenza richiede operazioni onerose in termini di tempo, in più la selezione risulta talmente semplice che l'utente la può svolgere in modo autonomo. Bisogna notare che l'HRTF selezionata con questo metodo non garantisce le stesse performance della propria HRTF, tuttavia offre prestazioni migliori di una HRTF generica misurata ad esempio su un manichino.

Alcuni esempi di questa tecnica sono:

- DOMISO [20]: la scelta viene effettuata tramite ripetute prove di ascolto di HRTF diverse ed esclusione delle peggiori, finché non si rimane con la migliore.
- selezione a due passi [21]: è divisa in due passi, nel primo si effettua una selezione ridotta di HRTF escludendo le peggiori tramite una prova di ascolto, nel secondo si sceglie da questa selezione a miglior corrispondenza fra il soggetto esterno ed il database di HRTF.
- corrispondenza tra misurazioni antropometriche [22]: si basa sul miglior riscontro della HRTF nel dominio antropometrico.

- tracciamento dei profili del padiglione auricolare [7]: si tracciano i contorni dell'orecchio e si applica un modello riflessivo in base ai raggi descritti.

2.3 ITD/ILD

Analizziamo più approfonditamente in questa sezione i concetti di ITD e ILD precedentemente introdotti.

2.3.1 Interaural time difference

Per quasi tutte le posizioni spaziali di una sorgente nello spazio, la propagazione sonora di queste sorgenti arrivano prima ad un orecchio che all'altro. Questo fenomeno fisico si traduce in una differenza di tempo interaurale, che prende il nome di interaural time difference (ITD). Questa differenza ha il suo minimo per le posizioni frontali, aumenta sistematicamente per le direzioni laterali, raggiunge il suo massimo. La posizione del massimo laterale dipende dalla posizione dell'orecchio, e il ritardo massimo tra le orecchie è di circa 0.7 ms per la testa di un adulto.

L'ITD è causato non solo dalla distanza tra le due orecchie, ma anche dalla diffrazione e rifrazione della testa. Questi effetti dipendono dalla frequenza, mentre è possibile approssimare l'ITD in tre diverse gamme di frequenza dalla frequenza più bassa alla più alta [23]: a frequenze inferiori a 2 kHz, la testa è la causa principale degli effetti d'ombra e quindi del ritardo dell'onda all'arrivo all'orecchio più distante. Con l'aumentare della frequenza la diffrazione causata dalla testa aumenta per il fatto che la lunghezza d'onda è più piccola rispetto alle dimensioni della testa. Di conseguenza, le onde iniziano a aggirare la testa, ma, l'influenza di queste onde sull'ITD, è limitata perchè non c'è quasi nessuna differenza di pressione sonora alle due orecchie [24]. Al di sopra di questa frequenza, il sistema uditivo elabora le informazioni nel dominio del tempo, e queste giocano un ruolo secondario per la localizzazione. Inoltre, a livelli più elevati di le frequenze, l'ILD diventa più importante e consente una migliore localizzazione del suono [25].

Per calcolare l'ITD sono state sviluppate varie soluzioni. Un metodo possibile è quella di calcolarlo sulla base di modelli fisici, con formule basate sulla velocità del suono nello spazio. Si va quindi da modelli più semplici come quello in [26] a modelli più complessi che tirano in ballo forme geometriche della testa più complesse [9], oppure la traslazione delle orecchie [8]. Un secondo metodo consiste nel estrarre la differenza dei tempi dalla risposta impulsiva. Anche in questo caso vi sono vari modi per

procedere.

I metodi di stima per l'ITD possono essere generalmente suddivisi in tre categorie:

- onset: sono i metodi basati nella rilevazione della differenza di tempo tra l'arrivo del segnale ad uno dei due orecchi e al suo opposto. I tempi di ingresso sono calcolati al superamento di un determinato livello di soglia [27].
- cross-correlation: sono i metodi basati sulla relazione per la quale il ritardo temporale massimizza la coerenza tra i segnali che raggiunge le due orecchie di un individuo [28].
- group delay: sono i metodi basati sull'analisi e la stima delle fasi gruppi di ritardo per determinare l'ITD [29].

2.3.2 Interaural level difference

L'ILD può essere determinato in base alla direzione e alla frequenza in funzione delle HRTF dell'orecchio destro e sinistro secondo la formula seguente:

$$ILD = 20 \log_{10} \left(\frac{HRTF_L}{HRTF_R} \right) \quad (2.2)$$

Questa differenza tra le orecchie è generato principalmente dal effetto che la testa stessa causa "nascondendo" l'orecchio al controlaterale. Più piccola è la lunghezze d'onda, più grande è l'ILD a causa di questo effetto. Tuttavia, a frequenze maggiori di 2kHz, questa non è l'unica ragione che influenza l'ILD, sono ad esempio presenti anche le onde che transitano attorno alla testa. L'ILD può essere stimato tramite il modello sferico della testa sotto le frequenze di 1kHz, dove questo modello risulta molto valido nella stima dell'ILD reale, come accade nel calcolo del ITD. Purtroppo questo modello non tiene conto dell'asimmetria della testa, della posizione delle orecchie e dei riflessi generati dalle spalle. In più va ricordato che l'udito umano non fa uso dell'ILD per frequenze inferiori a 2KHz [25]. Quindi piuttosto che modellare l'ILD, è preferibile calcolare l'HRTF, dato che gli effetti prodotti dall'ILD sono contenuti indirettamente nella funzione di trasferimento.

2.4 HRTF database

La definizione generale definisce un database come un archivio di dati strutturato in modo da razionalizzare la gestione e l'aggiornamento delle informazioni e da permettere lo svolgimento di ricerche complesse, e anche un database di HRTF si attiene a questa definizione. Nel nostro caso i database immagazzinano informazioni riguardanti i dati estratti attraverso l'estrazione di HRTF vista in precedenza. Attraverso questi file, è possibile operare con le varie applicazioni è utilizzare i dati raccolti in svariati ambiti, ad esempio è possibile utilizzarli per confrontarli con i parametri dell'utente per selezionare l'HRTF più simile a quella individuale della persona, oppure caricarla in un ambiente virtuale per aumentare il senso di immersività e quindi il realismo.

La maggior parte dei database esistenti sono memorizzate nel formato Matlab, cioè usano una convenzione di file Matlab per memorizzare matrici predefinite. Un'alternativa è SDIF, un formato generale per la memorizzazione di dati audio, che è stato adattato alle HRTF consentendo di memorizzare i dati di un singolo ascoltatore in una rappresentazione mista basata su file di testo e file binari. Un ulteriore metodo di memorizzazione sono i file ".wav" compressi. Questi sono un altro esempio di formato binario misto, che richiede inoltre una descrizione in file di testo separati per poter interpretare i dati [30].

2.4.1 CIPIC

Il database CIPIC è alla base del lavoro di questa tesi, perciò merita un'analisi più approfondita. Il CIPIC Interface Laboratory dell'Università della California, ha rilasciato un database di HRTF di pubblico dominio contenente i dati di 45 soggetti analizzati. Questo oltre a contenere 1250 risposte all'impulso per ognuna delle orecchie dei soggetti, include una serie di misure antropometriche che possono essere utilizzate per studi di vario genere. Tutte le HRTF sono state misurate con il soggetto seduto al centro di un cerchio di 1 m di raggio il cui asse è stato allineato con il proprio l'asse interurale, mentre la posizione della testa del soggetto non era vincolata. Attraverso tecniche appropriate sono state rimossi gli effetti di disturbo dovuti alle riflessioni nella stanza. La lunghezza di ogni HRIR è di 200 campioni, corrispondente ad una durata di circa 4,5 ms. La posizione della sorgente sonora è stata specificata dall'angolo azimutale θ e dall'angolo di elevazione ϕ in coordinate interaurali. L'elevazione è

stata campionata uniformemente in 50 valori, con variazione di 5.625° , da -45° a $+230.625^\circ$. Per l'azimuth si è optato per un campionamento che parte a -80° , -65° , -55° , da -45° a $+45^\circ$ con un passo di 5° , per finire con i valori $+55^\circ$, $+65^\circ$ e $+80^\circ$. Questo campionamento è mostrato in Fig.2.6.

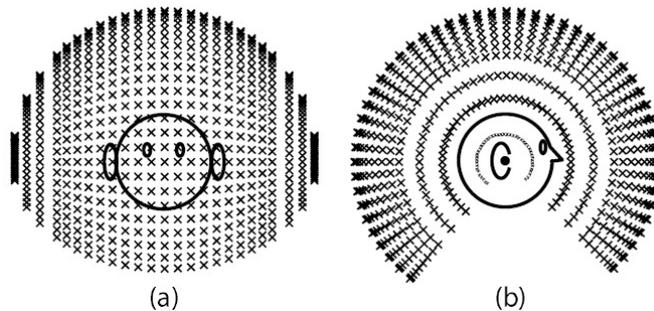


Figura 2.6: Punti di misurazione del database CIPIC. [6]

Per quanto riguarda le misure antropometriche, sono state acquisite 27 misure, 17 per la testa e il busto Fig.2.7 e 10 per l'orecchio Fig.2.8. La maggior parte dei parametri antropometrici sono stati estratti da fotografie digitali ad alta risoluzione. I parametri di altezza e circonferenza x_{14} ; x_{15} ; x_{16} ; x_{17} sono stati misurati con un nastro adesivo. La profondità della conca, d_8 e l'angolo del pinna, θ_2 , sono stati misurati con uno stilo per la digitalizzazione Polhemus 3D.

2.4.2 Spatially Oriented Format for Acoustics

Spatially Oriented Format for Acoustics(SOFA) è un formato di file per la memorizzazione di dati acustici ad orientamento spaziale come HRTFs o le directional room impulse responses (BRIRs, DRIRs). Questo formato è stato standardizzato nel 2015 dalla Audio Engineering Society(AES). SOFA ha lo scopo di rappresentare i dati spaziali in modo generale, permettendo di memorizzare non solo HRTF ma anche dati più complessi.

Gli obiettivi di SOFA sono:

- l'utilizzo di schemi di misura non arbitrari, che quindi, non sono limitati a griglie di punti regolari o con distanze costanti;
- dati auto descritti con una rigorosa definizione, dove tutte le informazioni delle misure sono inserite nel file;

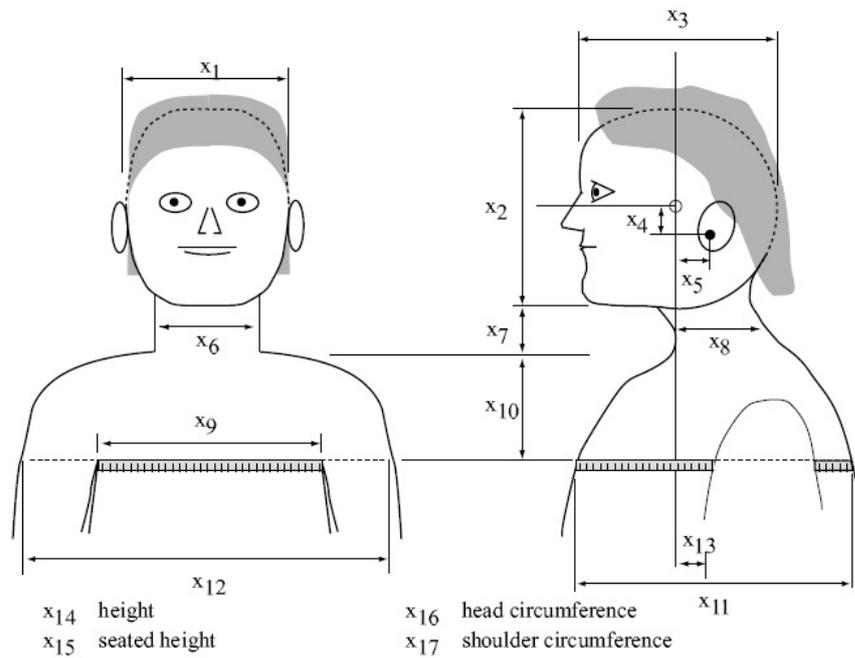


Figura 2.7: Misure di testa e torso. [6]

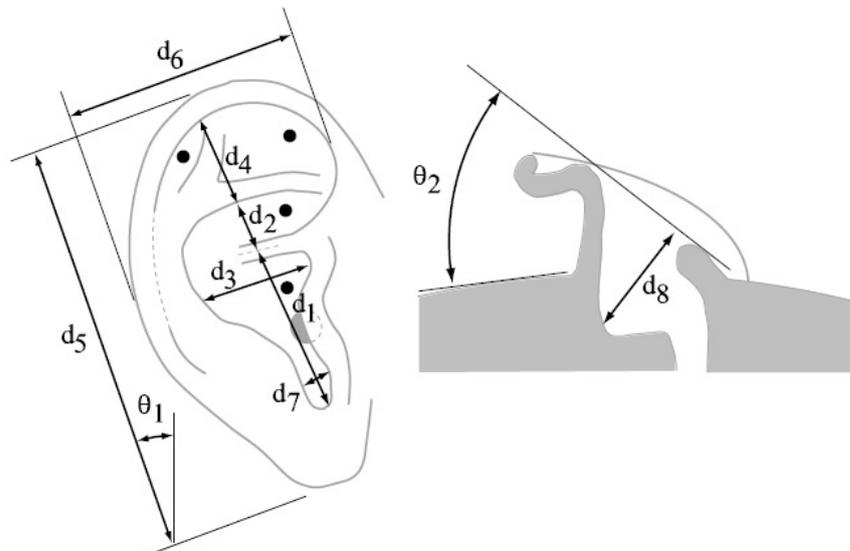


Figura 2.8: Misure dell'orecchio. [6]

- supporto di file parziale e di rete;
- utilizzo di file binari compressi per un'efficiente archiviazione e trasferimento.

Le specifiche SOFA mirano a soddisfare tutti questi requisiti. Le informazioni sono

memorizzate in un file binario basato su netCDF-4, e una descrizione più approfondita delle specifiche è disponibile in [31].

I file SOFA¹ possono essere letti e modificati da software e da interfacce API. Tra queste quelle usate in questo lavoro sono: HDF5View², che permette la consultazione dei dati numerici del file e SOFA API for Matlab/Octave³, che permette la creazione, attraverso matlab, di file ".SOFA" a partire dai dati di altri database salvati in estensione Matlab.

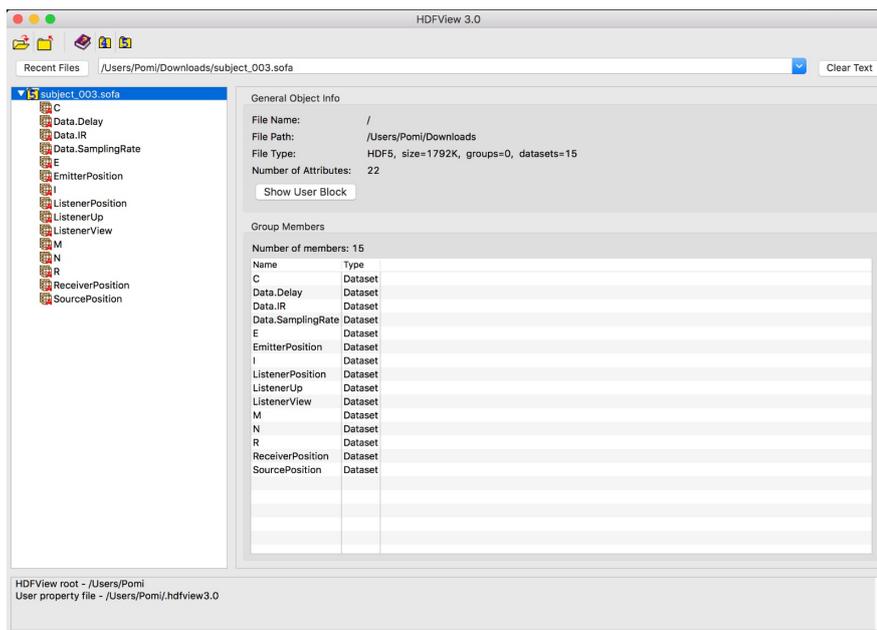


Figura 2.9: Esempio di file SOFA aperto utilizzando HDF5View. Tra i vari parametri troviamo Data.IR che racchiude la risposta impulsiva e SourcePosition che descrive le posizioni spaziali dei relativi punti.

¹[https://www.sofaconventions.org/mediawiki/index.php/SOFA_\(Spatially_Oriented_Format_for_Acoustics\)](https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics))

²<https://www.unidata.ucar.edu/software/netcdf>

³<https://sourceforge.net/projects/sofacoustics/>

3

Background e stato dell'arte

Dopo aver introdotto tutta la base teorica di questa tesi possiamo, in questo capitolo, descrivere il lavoro d'implementazione svolto, andando ad analizzare nel dettaglio l'obiettivo che ci siamo proposti e la bibliografia collegata.

3.1 Obiettivi della tesi

L'idea generale dietro questa tesi è la selezione, da un database, dell'ITD, tramite misure antropometriche, che più si avvicina a quello individuale dell'utente. La base di questo lavoro è descritta in [10], dove viene scelta la HRTF personalizzata più simile a quella individuale dell'utente. Si è deciso quindi di ampliare questa interfaccia con la possibilità di ottenere, oltre all'HRTF migliore, anche ITD migliore. Per fare questo si estraggono i dati antropometrici necessari dall'utente, questi vengono poi elaborati attraverso opportuni modelli e si va a trovare il soggetto con l'ITD più simile. Come ultima operazione i file dei due soggetti, quelli delle HRTF e ITD più simili, vengono combinati assieme per ottenere una miglior spazializzazione.

3.2 Selezione di HRTF basata sui contorni

Come detto prima, la base di questa tesi deriva da un lavoro precedentemente svolto [16] [10]. Andiamo quindi ad analizzarlo nel dettaglio.

Lo studio propone una tecnica di selezione di HRTF basata sul uso di immagini, ed è determinata dalla profonda relazione tra la forma della dell'orecchio esterno e i notches della funzione di trasferimento [7]. Utilizzando un'immagine 2D del padiglione del soggetto, la procedura seleziona da un database il set di HRTF che meglio si adatta alla sua antropometria. Questo strumento permette quindi la scelta di HRTF personalizzate anche ad utenti meno esperti in questo campo, dato che basta l'uso di una foto e il tracciamento di misure. L'idea principale è quella di disegnare i contorni dell'orecchio su un'immagine. Le distanze dall'ingresso del canale uditivo definiscono le riflessioni generate sui bordi del padiglione auricolare e queste portano a dei notches nello spettro delle HRTF. Di conseguenza, è possibile utilizzare tali dati antropometrici e i corrispondenti parametri spettrali per scegliere la migliore corrispondenza nel database di HRTF disponibili, dopo che anche queste sono state analizzate.

3.2.1 Mismatch in frequenza

L'estrazione di HRTF basata sulle riflessioni e sui contorni si avvale di una descrizione approssimativa degli effetti acustici dell'orecchio sui suoni in ingresso. In particolare, la distanza d_c tra un punto di riflessione e l'ingresso del condotto uditivo è data da:

$$d_c(\phi) = \frac{ct_d(\phi)}{2} \quad (3.1)$$

dove $t_d(\phi)$ è il ritardo temporale, dipendente dall'elevazione, tra l'onda diretta e quella riflessa mentre c è la velocità del suono in condizioni standard.

La frequenza del notch corrispondente dipende dal segno della riflessione. Supponendo che il coefficiente di riflessione sia positivo, si genera un notch a tutte quelle frequenze dove la differenza di fase tra l'onda riflessa e l'onda diretta è uguale a π :

$$f_n(\phi) = \frac{2n+1}{2t_d(\phi)} = \frac{c(2n+1)}{4d_c(\phi)} \quad (3.2)$$

dove $n \in \mathbb{N}$. La prima notch si trova quando $n = 0$, ottenendo il seguente risultato:

$$f_0(\phi) = \frac{c}{4d_c(\phi)} \quad (3.3)$$

Se ora prendiamo in esame il database CIPIC otteniamo che la maggior parte degli utenti presenta riflessione con coefficiente negativo nella HRIR [32]. Con questa assunzione si può creare un modello più accurato, dove i notch si trovano ad una differenza di fase di:

$$f_n(\phi) = \frac{n+1}{t_a(\phi)} = \frac{c(n+1)}{2d_c(\phi)} \quad (3.4)$$

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

$$f_0(\phi) = \frac{c}{2d_c(\phi)} \quad (3.5)$$

Dopo aver visto come calcolare le riflessioni, parliamo adesso del padiglione auricolare e dei suoi profili. Risulta possibile identificare 5 differenti profili [7]:

- C_1 : bordo interno dell'elice
- C_2 : bordo esterno dell'elice
- C_3 : bordo esterno della conca
- C_4 : antielice e bordo inferiore della conca, praticamente in mezzo al padiglione
- C_5 : radice dell'elice e superficie sottostante

Di queste, le prime tre sono legate ad altrettanti notch di frequenza [7], e quindi il modello può essere limitato a questi, diventando quindi:

- primo contorno: helix border C_1
- secondo contorno: antihelix and concha inner wall C_2
- terzo contorno: concha outer border C_3

Definiamo ora una funzione di mismatch, cioè una funzione che identifica la diversità tra le due parti, tra due HRTF come:

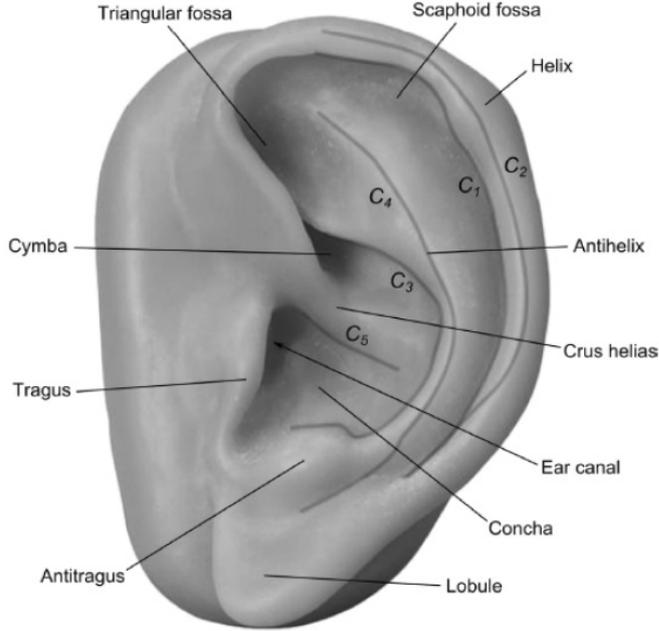


Figura 3.1: Rappresentazione dei cinque contorni principali. [7]

$$m = \frac{1}{n} \sum_{i=1}^n \frac{\omega_i}{N_\phi} \sum_{\phi} \frac{|f_0^i(\phi) - F_0^i(\phi)|}{F_0^i(\phi)} \quad (3.6)$$

dove n è il numero di contorni considerati, ω_i è una combinazione convessa dei pesi di ogni C_i , ϕ è l'angolo di elevazione dov'è situato il corrispondente notch in frequenza del HRTF mentre $N_\phi \in \{1, \dots, n\}$ è il numero dei campioni considerati per ogni differente elevazione per normalizzare la sommatoria, quando $f_0^i(\phi) \neq 0 \wedge F_0^i(\phi) \neq 0$. $f_0^i(\phi)$ rappresenta le frequenze estratte dai contorni e $F_0^i(\phi)$ quelle dal database di HRTF.

Considerando quindi i tre contorni presi in esame la formula diventa:

$$m_{(k,n)} = \sum_{i=1}^3 \frac{\omega_i}{N_\phi} \sum_{\phi} \frac{|f_0^{i(k,n)}(\phi) - F_0^i(\phi)|}{F_0^i(\phi)} \quad (3.7)$$

dove (k, n) con $(0 \leq k < K)$ e $(0 \leq n < N)$ si riferisce ad una particolare coppia composta da punto focale e canale uditivo, $i = 1; 2; 3$ sono i contorni di riferimento e ω_i è una combinazione convessa dei pesi di ogni C_i tale che $\sum_i \omega_i = 1$

Le metriche considerate sono in tutto 10, e sono:

- contorno C_1 ($\omega_1 = 1; \omega_2 = \omega_3 = 0$) con coefficiente di riflessione negativo.
- contorno C_1 ($\omega_1 = 1; \omega_2 = \omega_3 = 0$) con coefficiente di riflessione positivo.
- contorno C_2 ($\omega_2 = 1; \omega_1 = \omega_3 = 0$) con coefficiente di riflessione negativo.
- contorno C_2 ($\omega_2 = 1; \omega_1 = \omega_3 = 0$) con coefficiente di riflessione positivo.
- contorno C_3 ($\omega_3 = 1; \omega_1 = \omega_2 = 0$) con coefficiente di riflessione negativo.
- contorno C_3 ($\omega_3 = 1; \omega_1 = \omega_2 = 0$) con coefficiente di riflessione positivo.
- contorno C_1, C_2, C_3 equipesati ($\omega_1 = \omega_2 = \omega_3 = 1/3$) con coefficiente di riflessione negativo.
- contorno C_1, C_2, C_3 equipesati ($\omega_1 = \omega_2 = \omega_3 = 1/3$) con coefficiente di riflessione positivo.
- contorno C_1, C_2, C_3 con pesatura ottimale ($\omega_1 = 0.66; \omega_2 = 0.24; \omega_3 = 0.1$) con coefficiente di riflessione negativo.
- contorno C_1, C_2, C_3 con pesatura ottimale ($\omega_1 = 0.66; \omega_2 = 0.24; \omega_3 = 0.1$) con coefficiente di riflessione positivo.

3.2.2 Mismatch antropometrico

Oltre al mismatch in frequenza appena descritto è stato considerato anche un mismatch che si basa sulle misure del padiglione auricolare.

Per prima cosa si estraggono le misure attraverso l'immagine dell'orecchio dell'utente. Le misure prese in esame sono:

- altezza esterna del padiglione: misurata dalla base del lobo fino al punto più "alto" dell'orecchio. In Fig.2.8 corrisponde a d_5 .
- altezza interna del padiglione: misura interna della conca dell'orecchio. In Fig.2.8 corrisponde a $d_1 + d_2 + d_4$.

A questo punto si procede con il calcolo della differenza tra le misure tracciate e quelle del database. Il confronto per la misura esterna verrà effettuato tra il valore d_5 presente nel database con la misura tracciata. Il valore di mismatch è calcolato con come:

$$m_{d_1} = |d_{ext} - d_5| \quad (3.8)$$

con d_{ext} misura tracciata. Il confronto per la misura interna avviene nel medesimo modo, sostituendo d_5 con la somma di d_1, d_2, d_4 . La formula diventa quindi:

$$m_{d_2} = |d_{int} - (d_1 + d_2 + d_4)| \quad (3.9)$$

con d_{int} misura tracciata.

3.2.3 Classificazione dei risultati

Per ordinare i vari mismatch vengono introdotte tre tipi diversi di metriche:

- average mismatch: sono i valori medi di mismatch, ordinati in ordine crescente, delle coppie (k, n) per ciascuna HRTF.
- average rank: ad ogni HRTF viene assegnato un punteggio legato ad un numero intero corrispondente alla sua posizione in classifica presa dai valori della partita precedente.
- top-M appearance: per un dato intero M, ad ogni HRTF, viene assegnato un punteggio in base al numero di volte in cui l'HRTF si colloca nelle prime posizioni M.

Utilizzando tutte e dieci le metriche, e valutando tutte le possibili classifiche si ottengono 30 risultati diversi per ogni HRTF.

I dati antropometrici sono invece ordinati secondo due classifiche:

- mismatch dell'altezza esterna: ordina secondo la formula di mismatch esterno.
- mismatch dell'altezza interna: ordina secondo la formula di mismatch interno.

In tutto quindi vengono create 32 classifiche.

Questo lavoro è stato implementato utilizzando l'applicazione Matlab, di cui una schermata può essere visualizzata in Fig.4.1. Nella sezione successiva analizzeremo la schermata e le sue funzioni.

3.3 Modelli per ITD

Come visto in precedenza esistono vari modi per il calcolo dell'ITD, è possibile ricavarlo tramite modelli o può essere estratto con differenti tecniche. Focalizziamoci ora sulle possibili modellazioni di ITD utilizzando delle similitudini della testa a figure geometriche semplici, per ricavare dei modelli per calcolare l'ITD. L'analisi di queste tecniche è molto importante e si pone alla base di questa tesi.

3.3.1 Modello sferico

Il modello sferico della testa, analizzato per la prima volta da Lord Rayleigh [33] è una buona prima approssimazione della testa umana, e la sua semplicità analitica facilita la progettazione di algoritmi per la spazializzazione del suono. Questo modello è generalmente utilizzato per predire o sintetizzare il suono spaziale, poiché permette di calcolare sia l'ITD che l'ILD utilizzando le dimensioni esatte del individuo. Misure errate portano alla generazione di errore di localizzazione.

Per ottenere l'ITD da questo modello possiamo utilizzare la formula di Woodworth [34] che nel caso in esame diventa:

$$ITD(\theta, a) = \begin{cases} \frac{a}{c}(\sin \theta + \theta), & \text{if } 0 \leq \theta \leq \frac{\pi}{2} \\ \frac{a}{c}(\pi - \sin \theta + \theta), & \text{if } \frac{\pi}{2} \leq \theta \leq \pi \end{cases} \quad (3.10)$$

Dove θ rappresenta l'angolo di azimuth, c è la velocità del suono e a è il raggio della testa. Serve quindi un metodo per determinare il raggio a partendo dalle misure antropometriche della testa che si ricava utilizzando metà delle misure x_1, x_2 , e x_3 della Fig.2.7 attraverso formule empiriche, stimate statisticamente sulla popolazione. Un modello lineare generale che può essere facilmente ottimizzato è quello di stimare il raggio della testa come somma ponderata delle dimensioni della testa, vale a dire,

$$a_e = \omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3 + b \quad (3.11)$$

Nella regressione sui dati eseguita in [35] si ottengono i valori: $\omega_1 = 0.51$, $\omega_2 = 0.019$, $\omega_3 = 0.18$ e $b = 32\text{mm}$. In questo caso quindi la formula per da utilizzare per

calcolare il raggio ottimo diventa:

$$a_e = 0.51X_1 + 0.019X_2 + 0.18X_3 + 32 \quad (3.12)$$

3.3.2 Modello sferico con la traslazione delle orecchie

Nel modello sferico è comune collocare le orecchie in posizioni diametralmente opposte e, come tali, questi modelli prevedono valori di ITD costanti su coni di confusione centrati sull'asse interaurale. In questo modo si perdono le informazioni dell'elevazione osservata nelle misurazioni ITD attraverso modelli non sferici della testa e con lo spostamento delle orecchie [9]. Quindi andiamo a considerare anche lo spostamento delle orecchie nel modello sferico per migliorarne l'approssimazione. Per prima cosa dobbiamo riutilizzare la formula di Woodworth and Schlosberg [34], che in assenza di semplificazioni dovute al posizionamento delle orecchie diviene:

$$\Delta T(\gamma, a) = \begin{cases} -\frac{a}{c}(\cos \gamma), & \text{if } 0 \leq |\gamma| < \frac{\pi}{2} \\ \frac{a}{c}[|\gamma - \frac{\pi}{2}|], & \text{if } \frac{\pi}{2} \leq |\gamma| < \pi \end{cases} \quad (3.13)$$

dove a è il raggio della testa. Questa formula permette di stimare la differenza di tempo di arrivo ΔT tra un punto di osservazione sulla sfera e il centro della sfera nello spazio libero. Da questo si può quindi ricavare l'ITD come differenza nei tempi di arrivo ΔT tra orecchio destro e sinistro o viceversa. L'angolo γ è definito attraverso il prodotto scalare tra il vettore \vec{e} e il vettore \vec{S} . Questi due vettori, nel caso in cui le orecchie non sono poste parallelamente, sono uguali a:

$$\vec{S} = \begin{bmatrix} \sin \Theta \\ \cos \Theta \cos \Phi \\ \cos \Theta \sin \Phi \end{bmatrix}, \quad \vec{e} = \begin{bmatrix} \pm \sqrt{a^2 - e_d^2 - e_b^2} \\ 0 \\ 0 \end{bmatrix} \quad (3.14)$$

dove (Θ, Φ) sono le direzioni del punto da calcolare e e_b, e_d rappresentano lo spostamento delle orecchie. Per calcolare γ ponendo $x_0 = [0, e_b, e_d]$ si utilizza:

$$\gamma = \cos^{-1} \left[\frac{\vec{S} \cdot \vec{e}}{\|\vec{S}\| \cdot \|\vec{e}\|} \right] = \cos^{-1} \left[\frac{(\vec{S}' - \vec{x}_0) \cdot (\vec{e}' - \vec{x}_0)}{\|\vec{S}' - \vec{x}_0\| \cdot \|\vec{e}' - \vec{x}_0\|} \right] \quad (3.15)$$

Quando le orecchie non sono più nella coordinata di $90^\circ, -90^\circ$ di azimuth, l'asse

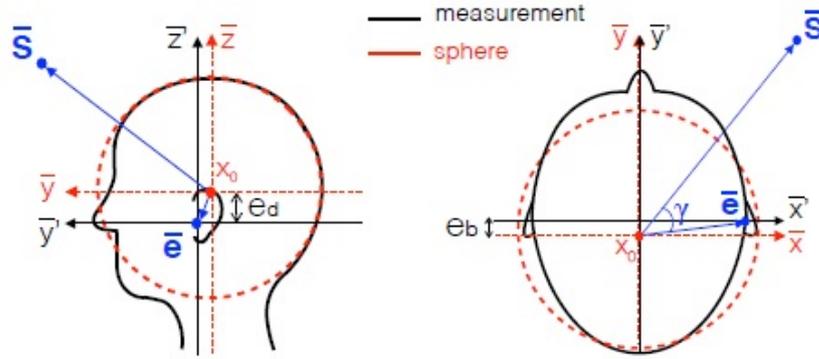


Figura 3.2: Rappresentazione delle orecchie spostate. [8]

X del sistema di coordinate non coincide più con quello interaurale. Lo spostamento delle orecchie all'indietro, definito con e_b , e verso il basso, definito con e_d , descrive la posizione delle orecchie nel sistema sferico, oppure, definisce lo spostamento del sistema di riferimento da quello interaurale, questi due effetti sono visibili in Fig.3.2. L'effetto dello spostamento all'indietro porta ad asimmetrie nell'ITD [8], i percorsi frontali risultano maggiori e quelli posteriori minori. L'ITD massimo è leggermente ridotto e spostato nella direzione che si oppone direttamente all'orecchio opposto. Lo spostamento verso il basso genera simmetrie in altezza, spostando le orecchie verso il basso, la lunghezza del percorso per le sorgenti elevate aumenta ed è di nuovo maggiore per l'orecchio controlaterale.

Poiché la testa umana non è una sfera perfetta, determinare il raggio della testa da utilizzare nella formula di Woodworth non è banale. Dobbiamo quindi ricavare una formula per stimare il raggio ottimo, come fatto in precedenza, tenendo conto anche dello spostamento delle orecchie. Prendendo le prime 5 misure del CIPIC, mostrate in Fig.2.7, e utilizzando un modello di regressione [8], e non considerando lo spostamento, si ottiene:

$$\hat{a}_{opt} = 0.44 \cdot X_1 + 0.23 \cdot X_3 + 3.2 \quad (3.16)$$

Un risultato in linea con la regressione vista in precedenza ad opera i Algazi et al.[35].

Considerando ora lo spostamento, e prendendo valori ottimo estratti dal database CIPIC [8], si ottiene che $e_{b_{opt}} = -0.94 \pm 0.5$ cm e $e_{d_{opt}} = 2.10 \pm 0.4$ cm. Applicando

questi dati al modello precedente si ottiene una formula di regressione di:

$$\hat{\alpha}_{opt} = 0.41 \cdot X_1 + 0.22 \cdot X_3 + 3.7 \quad (3.17)$$

Sia in questo caso, che nel precedente, X_2 non è stato considerato perché insignificante, mentre X_4 e X_5 sono risultati statisticamente non rilevanti nello spostamento [8]. Un confronto usando la formula di Woodworth per ricavare l'ITD tra questa regressione e i parametri reali estratti dal CIPIC si può vedere in Tabella 3.1.

		Spherical Antipodal Ears	Spherical Ear Shifts	Ellipsoidal Ear Shifts
Direct	Abs. error	34.79	29.82	39.43
	RMS error	41.89	35.85	48.92
Optimized	Abs. error	25.84	21.21	20.00
	RMS error	31.31	25.28	24.12
Predicted	Abs. error	26.67	22.84	22.33
	RMS error	32.28	27.15	26.71

Tabella 3.1: Confronto attraverso media assoluta e media quadratica tra il modello sferico con e senza traslazione e quello ellittico per i parametri diretti, predetti e ottimi. [8]

Il confronto evidenzia come i due risultati siano molto simili, validando l'uso di questa formula per approssimare il raggio ottimo se confrontato con quello reale.

Analizzando il modello con orecchie traslate sul piano orizzontale, otteniamo la Fig. 3.3.

Le linee rosse indicano il limite dell'impercettibilità dell'errore dell'ITD [36], cioè, quando vengono sorpassate l'effetto dell'ITD sulla percezione del suono è non trascurabile. Le linee grige e nere rappresentano rispettivamente, l'errore per ogni soggetto analizzato e l'errore medio. Questo modello sovrastima l'ITD fino ad un angolo laterale di $\theta = 60$ e lo sottostima per $\theta > 60$. Questo errore è probabilmente dovuto alla semplificazione di testa come sfera, anche se la media dei valori non sorpassa comunque i limiti di udibilità.

3.3.3 Modello ellissoidale

Oltre al modello sferico possiamo approssimare la testa come un ellissoide, questo modello prende il nome di modello ellissoidale della testa. A differenza del modello sferico, in questo modello deve essere considerata anche l'elevazione per quanto ri-

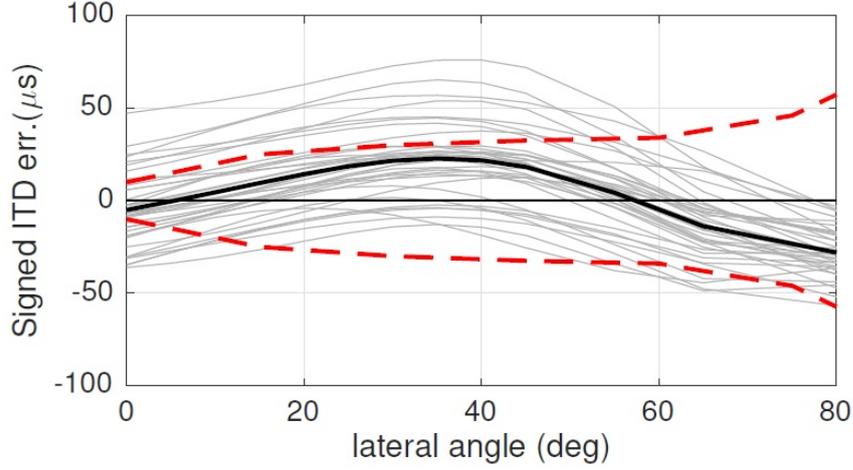


Figura 3.3: Performance del modello sferico con orecchie spostate. Le linee rosse rappresentano l'udibilità dell'errore, quelle grige l'errore dei soggetti, e quella nera la media degli errori. [8]

guarda il calcolo dell'ITD, questo è dovuto al fatto che se la testa non è sferica le orecchie non sono locate ai bordi di una circonferenza e questo fa sì che il percorso delle onde dalla sorgente all'orecchio vari anche con l'elevazione. Per esprimere questa dipendenza è possibile utilizzare un ellissoide [9]. Per enunciare il modello analitico prendiamo $\bar{s} = [s_1, s_2, s_3]^T$ il vettore origine del suono e scriviamo $\bar{e} = [e_1, e_2, e_3]^T$ il vettore dell'orecchio opposto. L'ellissoide è definito da:

$$(\bar{x} - \bar{x}_0)^T A (\bar{x} - \bar{x}_0) = 1 \quad (3.18)$$

dove $\bar{x} = [x_1, x_2, x_3]^T$ è un punto nell'ellissoide, \bar{x}_0 ne rappresenta il centro e A è una matrice 3×3 , tipicamente definita come

$$A = \begin{vmatrix} 1/a_1^2 & 0 & 0 \\ 0 & 1/a_2^2 & 0 \\ 0 & 0 & 1/a_3^2 \end{vmatrix} \quad (3.19)$$

con a_k lunghezza dell'asse principale lungo la direzione della coordinata k -esima. Come già detto in precedenza e_b è lo spostamento dell'orecchio verso la parte posteriore della testa e e_d è lo spostamento verso il basso. In questo caso, $x_0 = [0, e_b, e_d]^T$ e $e_1 = \pm a_1 \sqrt{1 - (e_b/a_2)^2 - (e_d/a_3)^2}$. Tutti i parametri sono raffigurati in Fig.3.4.

Riassumendo il modello è descritto da 5 misure: metà larghezza a_1 , l'altezza a_2 e la profondità a_3 dell'ellissoide e le posizioni e_b ed e_d , relative allo spostamento delle

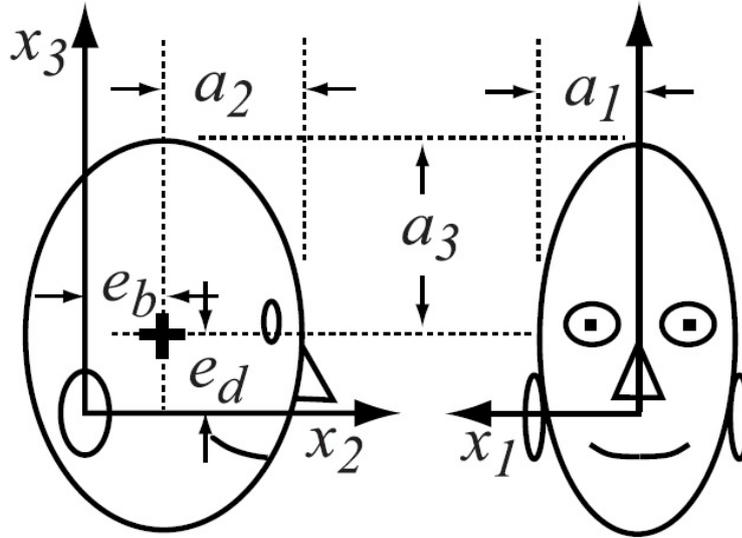


Figura 3.4: Dimensioni della testa e posizione delle orecchie. [9]

orecchie. Per calcolare l'ITD viene utilizzata la tecnica descritta in [9]. In questo modello vengono considerate le onde che percorrono i percorsi attorno all'ellissoide, valide solo ad alte frequenze. La lunghezza del percorso all'orecchio ipsilaterale è calcolata come distanza diretta, e quella controlaterale è calcolata come lunghezza diretta fino al punto di tangenza a cui viene sommata la lunghezza dell'arco più corto da questo punto all'orecchio.

Come visto per il modello precedente, possiamo utilizzare un'analisi di regressione per ottenere un modello per estrarre le misure per il calcolo dell'ITD partendo dalle misure del CIPIC:

$$\hat{a}_{1_{opt}} = 0.48 \cdot X_1 + 4.7\hat{a}_{3_{opt}} = 0.45 \cdot X_1 + 6.2 \quad (3.20)$$

Per le altre misure non sono state trovate corrispondenze statistiche, ma è possibile comunque utilizzare i valori ottimi medi che equivalgono a: $\bar{a}_{2_{opt}} = 11 \pm 0.5$ cm, $\bar{e}_{d_{opt}} = 2.9 \pm 0.5$ cm, $\bar{e}_{b_{opt}} = -1 \pm 0.6$ cm.

Calcolando l'ITD attraverso il modello con orecchie spostate, per tutti i soggetti del CIPIC nel piano orizzontale come fatto in precedenza per il modello sferico, otteniamo il grafico in Fig.3.5.

Anche in questo caso abbiamo una sovrastima e una sottostima dell'ITD per valori maggiori e minori di $\theta = 60$, ma in questo caso, l'errore di stima è leggermente ridot-

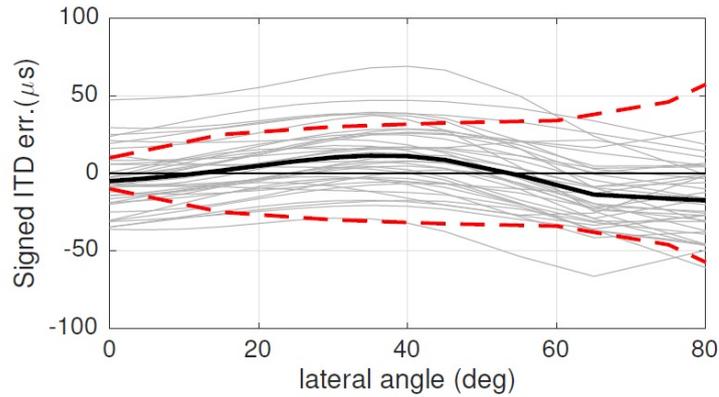


Figura 3.5: Dimensioni della testa e posizione delle orecchie. [8]

to. Va comunque riportato che la differenza generale dell'errore, descritto in Tab.3.1, tra i due modelli è molto lieve, implicando che ambedue i modelli offrono una buona stima nel calcolo dell'ITD.

In questa tesi si è scelto di utilizzare il modello sferico dato che è di più facile elaborazione computazionale dando comunque valori di approssimazione molto simili.

3.4 Tecniche di estrazione e modifica di ITD

Chiudiamo questo capitolo parlando di due tecniche essenziali usate in questo lavoro, l'estrazione dell'ITD dei soggetti del database, che saranno poi confrontati con quello calcolato e la manipolazione dell'ITD a fase minima che ci permetterà di elaborare una nostra versione del HRTF del soggetto.

3.4.1 Estrazione

Come già esplicitato in precedenza, esistono vari metodi per l'estrazione dell'ITD, qui ci concentriamo sul metodo di threshold e sulla cross-correlation perché questi saranno successivamente utilizzati in questa tesi.

Threshold

In questa tecnica, che si applica in funzione del tempo, viene confrontato l'onset, cioè il picco della funzione, delle HRIR destra e sinistra. Spesso questi picchi non sono rilevati al punto massimo, ma piuttosto diversi decibel al di sotto. Questo rende il rilevamento del livello di soglia più robusto, poiché l'HRIR possono avere diversi picchi vicino all'asse interaurale. L'ITD è calcolato come differenza di tempo tra l'onset dell'orecchio destro e sinistro secondo:

$$ITD_{threshold} = \tau_R - \tau_L \quad (3.21)$$

dove τ_R , τ_L sono gli onset calcolati dall'inizio del segnale al picco.

cross-correlation

In questo caso l'ITD viene calcolato nel dominio del tempo attraverso la cross-correlazione:

$$ITD_{cross-correlation} = argmax_{\tau}(HRIR_L \cdot HRIR_R) \quad (3.22)$$

Risultati

Un confronto tra i vari metodi è disponibile in [12] e visualizzabile in Tab.3.2 , dove sono stati confrontati diversi metodi di estrazione. Il risultati mostrano che il miglior metodo è quello di threshold con una soglia di -30dB applicata ad una versione filtrata con passa-basso a 3kHz dell'HRIR [12].

Rank number	Sum criteria (± 0.5 JND)		Sum criteria (± 1 JND)		Product criteria (± 0.5 JND)		Product criteria (± 1 JND)	
1	Threshold -30 dB lp	(0.43)	Threshold -30 dB lp	(0.71)	Threshold 30 dB lp	(1.00)	Threshold 30 dB lp	(1.00)
2	MaxIACCe lp	(0.39)	Threshold -20 dB lp	(0.66)	MaxIACCe lp	(0.39)	Threshold 20 dB lp	(0.57)
3	Threshold -20 dB lp	(0.38)	CenIACCr bb	(0.62)	CenIACCr lp	(0.33)	CenIACCr bb	(0.37)
4	CenIACCr lp	(0.37)	MaxIACCe lp	(0.61)	Threshold 20 dB lp	(0.29)	MaxIACCe lp	(0.34)
5	Cen-e ² lp	(0.34)	CenIACCe lp	(0.61)	Cen-e ² lp	(0.10)	CenIACCr lp	(0.33)

Tabella 3.2: Confronto tra i vari metodi di estrazione di ITD [12].

Combinazione di HRTF

La modifica dell'ITD di un'HRTF, andandolo a sostituire con un altro, può essere applicata, nel dominio delle frequenze, mediante la ricostruzione a fase minima di

un'HRTF più il delay di un'altra. In pratica si deve calcolare l'ITD da applicare, con uno dei metodi visti, e lo si aggiunge convertito in fase alle HRTF da manipolare portandola a fase minima.

Un'ulteriore ottimizzazione è quella di modificare la magnitudine per risolvere i problemi che affliggono l'HRTF a bassa frequenza, infatti, le HRTF misurate non sono affidabili alle basse frequenze a causa dell'uso di altoparlanti relativamente piccoli per le misure [37]. Per porre rimedi a questo problema si possono combinare le risposte in frequenza dell'HRTF scelta tramite ITD e dell'HRTF selezionata tramite i contorni dell'orecchio. Scriviamo quindi, H_h l'HRTF del modello con cui selezionare l'ITD e H_c l'HRTF basata sui contorni dell'orecchio da cui possiamo ricavare H_s l'HRTF combinata come:

$$A_s(\omega) = \begin{cases} A_h(\omega) & \text{if } \omega < \omega_l \\ A_h(\omega) + \frac{A_c(\omega) - A_h(\omega)}{\omega_h - \omega_l}(\omega - \omega_l) & \text{if } \omega_l < \omega < \omega_h \\ A_c(\omega) & \text{if } \omega > \omega_h \end{cases} \quad (3.23)$$

dove $A_s(\omega) = \log |H_s(\omega)|$, $A_h(\omega) = \log |H_h(\omega)|$, $A_c(\omega) = \log |H_c(\omega)|$, mentre per i valori di frequenza si ha: $\omega_l = 250\text{Hz}$, $\omega_h = 1000\text{Hz}$.

4

Implementazione del software di personalizzazione

Dopo aver parlato della parte teorica e della letteratura impiegata in questo lavoro, Presentiamo in questo capitolo come queste nozioni teoriche sono state sviluppate in pratica e come sono state aggiunte al precedente lavoro su cui si fonda questa tesi.

Come già detto precedentemente, lo scopo della tesi è di creare un software per la selezione delle HRTF personalizzata secondo le caratteristiche antropometriche dell'utente. Per fare ciò si è scelto di utilizzare matlab, una piattaforma che comprende un suo linguaggio di programmazione e che è principalmente usato per le elaborazioni numeriche e per il calcolo statistico, e in particolare, utilizzare questo programma per creare una finestra grafica per facilitare la personalizzazione delle opzioni da scegliere anche per utenti meno esperti.

4.1 Interfaccia grafica

In questo capitolo analizziamo l'interfaccia grafica(GUI), esaminando anche la precedente versione, puntualizzando le eventuali modifiche che sono state apportate.

In Fig.4.1 è presente la schermata della versione precedente alle modifiche [10], mentre, nella Fig.4.2 è presente la versione finale elaborata in questa tesi.

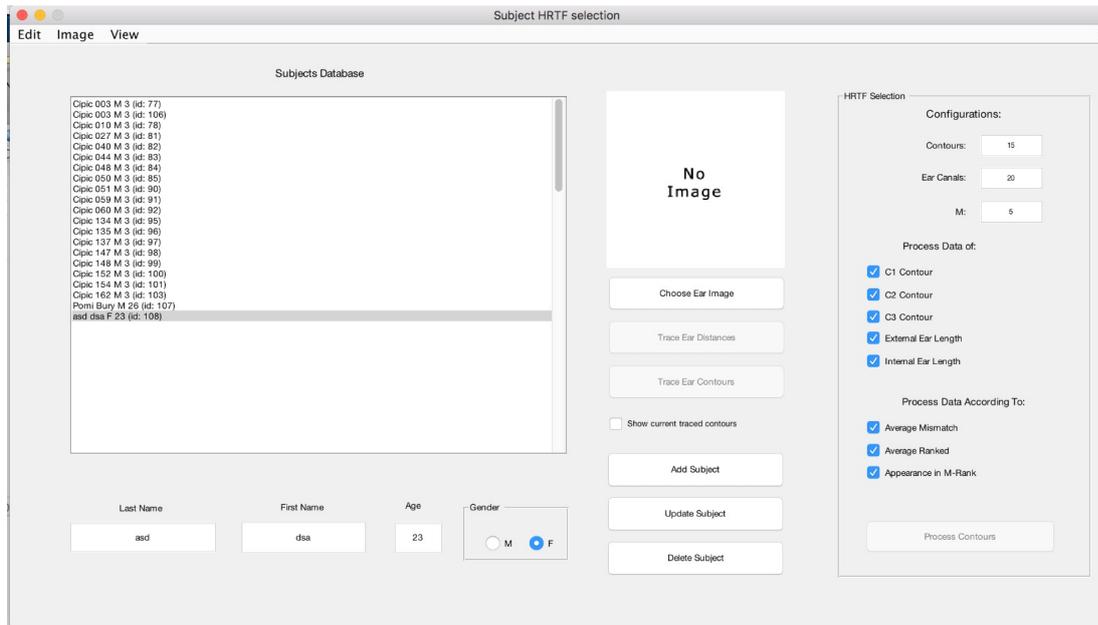


Figura 4.1: Schermata della GUI prima delle modifiche. [10]

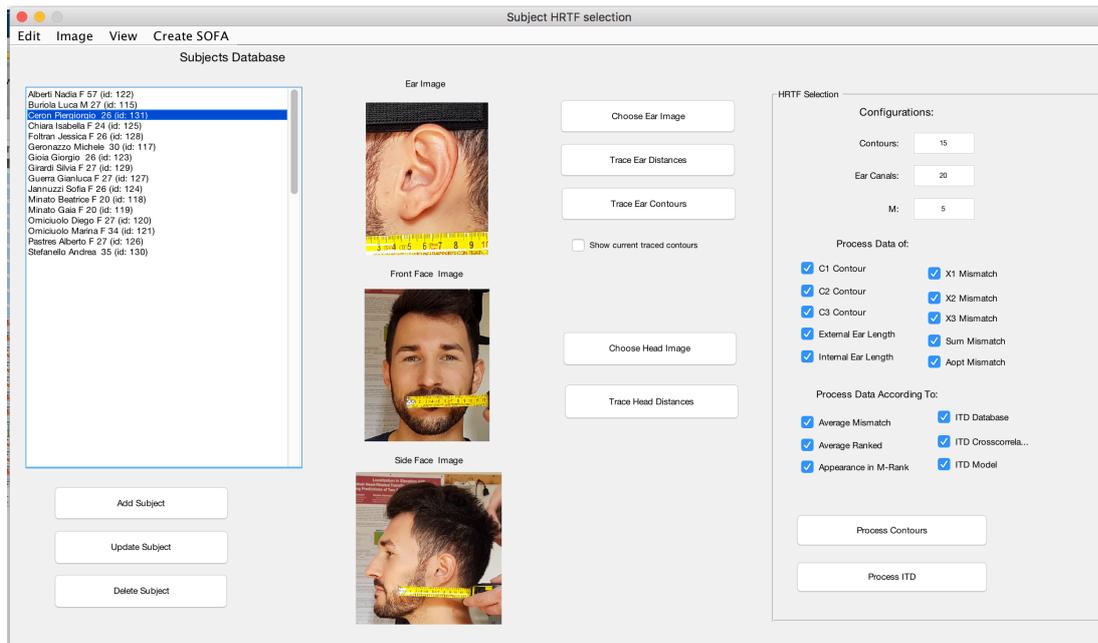


Figura 4.2: Schermata della GUI dopo delle modifiche.

4.1.1 Opzioni d'inserimento dati

Nella versione precedente erano presenti un box di testo dove venivano visualizzati le informazioni generali del soggetto: nome, cognome, sesso, età e ID del database.

Sotto a questa vi erano delle caselle di testo dove inserire i dati sopracitati, e in parte i tasti per aggiungere, cancellate o editare i soggetti. Infine era possibile inserire la foto dell'orecchio dell'utente per le elaborazione dei contorni tramite il tasto *Choose Ear Image*, e questa veniva visualizzata nella finestrella appena sopra.

Nella nuova versione è stato mantenuta la box per la visualizzazione dei dati, ma le opzioni per intervenire sugli utenti sono state spostate. Ora premendo il tasto *Add Subject* oppure *Update Subject* si apre un'ulteriore finestra, che permette la modifica o l'aggiunta di un soggetto, mentre per l'eliminazione basta utilizzare il tasto *Delete Subject*. L'immissione della foto dell'orecchio è stata mantenuta invariata, ma è stata spostata la sua posizione, e sono state aggiunte le immagini frontali e laterali del volto dell'utente, utilizzate per l'estrazione delle misure per il calcolo dell'ITD, oltre al relativo tasto per il loro inserimento chiamato *Choose Head Image*. Una volta premuto questo tasto si apre una finestra che invita all'immissioni delle due foto, e riporta un errore se le foto non sono caricate nella cartella "img" presente nei file del plugin.

4.1.2 Rilevamento dei dati per l'elaborazione

Nella parte centrale sono presenti i pulsanti per l'inserimento dei dati. Prima di poterli utilizzare, però, è necessario settare il *pixel to meters*. Questa opzione si trova nella tendina in alto all'interno dell'opzione *Image*, e permette di convertire le distanze prese in pixel nelle relative misure in centimetri. Per fare questo basta misurare un centimetro nella foto che viene visualizzata quando questa funzione viene eseguita. Esistono due diversi pixel to meters, uno per il tracciamento dei contorni e uno per il calcolo dell'ITD, dato che lavorano su fotografie diverse.

I tasti per ricavare le misure necessarie all'individuazione dell'HRTF sono il *Trace Ear Distances*, che estrae le misure dell'orecchio, e il *Trace Ear Contours*, che attraverso i parametri *Contours*, *Ear Canals* e *M* ricava i contorni C_1 , C_2 e C_3 . Il funzionamento di questi tasti è rimasto invariato, mentre è stato aggiunto il tasti per il calcolo dell'ITD: *Trace Head Distances*. Tramite questo tasto è possibile selezionare le misure antropometriche del volto, una volta premuto infatti, si aprono le immagini frontali e laterali del utente e viene chiesto di tracciare una linea per le 3 misure necessarie per i calcoli: x_1 , x_2 , x_3 .

4.2 Metriche di confronto

Parliamo ora delle metriche utilizzate per l'elaborazione dei confronti e per ottenere i risultati. Come già citato nel capitolo precedente [10], le opzioni di selezione dei contorni permettono di decidere quali dati includere nell'analisi, mentre, le tre caselle da spuntare al di sotto di questi permettono di scegliere in che modo elaborare i dati e quali risultati saranno stampati, secondo i metodi precedentemente spiegati. Per quanto riguarda le selezioni introdotte nella nuova versione, le scelte a disposizione sono 5:

- X_1 Mismatch
- X_2 Mismatch
- X_3 Mismatch
- Sum Mismatch
- A_{opt} Mismatch
- ITD Database
- ITD Cross-correlation
- ITD Model

Basato sulle misure antropometriche

I primi cinque confronti sono di tipo antropometrico. In questo caso le misure ricavate dal soggetto attraverso il tasto *Trace Head Distances*, vengono confrontate con quelle estratte dal database CIPIC. Nello specifico nei primi tre casi, X_1 , X_2 , X_3 , vengono comparate tra loro le singole misure e viene generata una classifica secondo la differenza minima assoluta, nel caso di Sum si utilizza la somma dell'errore assoluto sul singolo di queste tre. In A_{opt} invece le misure citate vengono combinate tre loro come nella formula di calcolo dell' A_{opt} 3.17 e viene quindi stilata una classifica in base alla differenza minima.

Basato sull'ITD

Negli ultimi 3 casi il confronto per la selezione del HRTF con l'ITD più simile al soggetto preso in esame, avviene in termini di confronto tra ITD. Si calcola attraverso l'equazione 3.13 il valore di ITD per ogni azimuth e per ogni elevazione, utilizzando come raggio della testa il valore di A_{opt} , e lo si confronta con quello nella medesima coordinata, ricavato in modo differente a seconda del metodo scelto, per ogni soggetto. Il risultato finale si ottiene utilizzando l'errore quadratico medio:

$$MSE = \frac{\sum_{i=1}^n (x(\theta, \phi) - \hat{x}(\theta, \phi))^2}{n} \quad (4.1)$$

dove $x(\theta, \phi)$ e $\hat{x}(\theta, \phi)$ rappresentano rispettivamente l'ITD di uno specifico soggetto del database e quello del nostro utente calcolata tramite il modello sferico con lo spostamento delle orecchie, mentre i varia per ogni punto dello spazio per un totale di 1250 volte. I valori di MSE ricavati vengono poi confrontati tra loro per determinare il migliore e viene stilata una classifica finale.

I 3 ITD scelti per il confronto sono stati precalcolati ed è sono stati inseriti in un file di tipo ".mat" per ogni utente. Questi contengono tutti i dati per ogni valore di coordinata, e sono stati poi raggruppati nella cartella ITD_Subject.

Nel dettaglio abbiamo:

- database: sono i valori di ITD già presenti nel database [6].
- cross-correlation: il metodo di estrazione in questo caso è la cross-correlation [12]. Come specificato nella sezione precedente è stato usato il metodo più valido per l'estrazione, cioè, una cross-correlation applicando un threshold con un livello di -30dB applicato ad una versione filtrata dell'HRIR con un filtro passa-basso a 3kHz. Per l'estrazione è stato usato lo script matlab "amtoolbox-full-0.9.9".
- model: in questo metodo L'ITD è stato calcolato basandoci sulle misure antropometriche e sul modello di Woodworth estesa 3.13, per considerare anche lo spostamento delle orecchie.

Per creare i file è stato scritto uno script matlab che genera in automatico tutti gli archivia ".mat" autonomamente.

4.3 Supporto SOFA

L'ultima importante modifica apporta riguarda la generazione dei file SOFA. A differenza della versione precedente, dove il risultato delle operazioni era solo di tipo grafico e i SOFA da utilizzare erano direttamente quelli del database CIPIC, qui è stata implementata la funzione di generazione dei file SOFA. Come già detto in precedenza infatti, si vuole generare un file che utilizzi le informazioni della pinna per la localizzazione verticale, e che utilizzi un ITD basato sulla similitudine antropometrica della testa per la localizzazione orizzontale. Questo si può fare utilizzando i contorni delle orecchie [10] per determinare quale soggetto sia il più simile all'utente. Una volta fatto ciò è possibile utilizzare l'HRIR è quindi l'HRTF di tale soggetto come base del file SOFA da creare, andando a modificare il suo ITD. Si sceglie quindi il soggetto più simile all'utente attraverso la metrica preferita e si estrae uno dei tre ITD pre-calcolati. A questo punto, basta inserire l'ITD selezionato nell'HRTF attraverso la tecnica a fase minima più delay spiegata in precedenza in 3.4.1

4.3.1 GUI per la creazione di SOFA

La funzione per la creazione si trova nella barra in alto all'interno dell'opzione Create SOFA. Permuta questa apparirà una finestra come mostrato in Fig.4.1. Nella parte sinistra è possibile selezionare i soggetti di origine del file. Quello sopra individua la selezione del soggetto dal database scelto tramite i confronti dei contorni dell'orecchio, mentre quella sotto identifica il soggetto del quale si vuole utilizzare l'ITD.

Al centro è presente al selezione del nome del file. Sono possibili tre scelte diverse:

- Selected Nome: attraverso questa opzione il nome sarà automaticamente generato partendo dalle informazioni dell'utente selezionato.
- Subject Name: attraverso questa opzione il nome sarà generato dall'unione dei soggetti scelti nei campi di selezione dei subject.
- Edit Name: selezionando questa opzione apparirà una cassella di testo dove sarà possibile inserire il nome del file.

Infine nella parte sinistra è presente la selezione dell'ITD da utilizzare. Qui infatti si decide quale dei 3 ITD del soggetto selezionato si vuole implementare nel file Sofa che verrà generato.

- Database: utilizza l'ITD estratto dal database CIPIC del soggetto che si è selezionato in ITD SELECTION.
- Cross Correlation: utilizza l'ITD ricavato tramite cross-correlation con i parametri visti prima del soggetto che si è selezionato in ITD SELECTION
- model: utilizza l'ITD calcolato tramite modello sferico del soggetto che si è selezionato in ITD SELECTION.



Figura 4.3: Schermata del menu' di generazione dei file Sofa. [10]

4.4 Uso del Plugin

Dopo aver descritto le modifiche apportare e il loro utilizzo, descriviamo infine come funziona il plugin dopo l'aggiornamento.

Per iniziare sono necessarie le foto dell'utente di cui si vuole ricavare l'HRTF personalizzata. Ne servono 3 da caricare nella cartella "img", una frontale, una laterale e una dell'orecchio. Nel fare le foto si deve controllare che il soggetto sia perfettamente perpendicolare alla camera, o la misura non sarà precisa e bisogna anche includere nella foto un metro, o un altro tipo di riferimento come termine di paragone per la successiva misura. Dopo aver caricato le foto nella cartella si può passare alla creazione dell'utente inserendo i dati negli appositi campi e selezionare le foto con i tasti appositi. A questo punto si imposta il pixel to meters di tutti e due i tipi di selezione. Per fare ciò si traccia una linea lunga quanto il riferimento precedentemente inserito, e si specifica poi la relativa lunghezza. Ora che il pixel to meters è stato regolato,

si possono usare tracciare le grandezze dell'orecchio e della testa, e si procede con il tracciamento dei contorni dell'orecchio. Una volta che tutte le variabili sono state rilevate, non rimane altro da fare che iniziare l'esecuzione delle procedure per il calcolo dei risultati. L'esito di questi sarà disponibile in grafici in formato ".pdf" nella cartella "results". L'ultima operazione che rimane è quella di creare il file Sofa, attraverso la funzione sopra descritta, dei soggetti che si decide di prendere in considerazione. Uno schema riassuntivo è rappresentato in figura 4.4

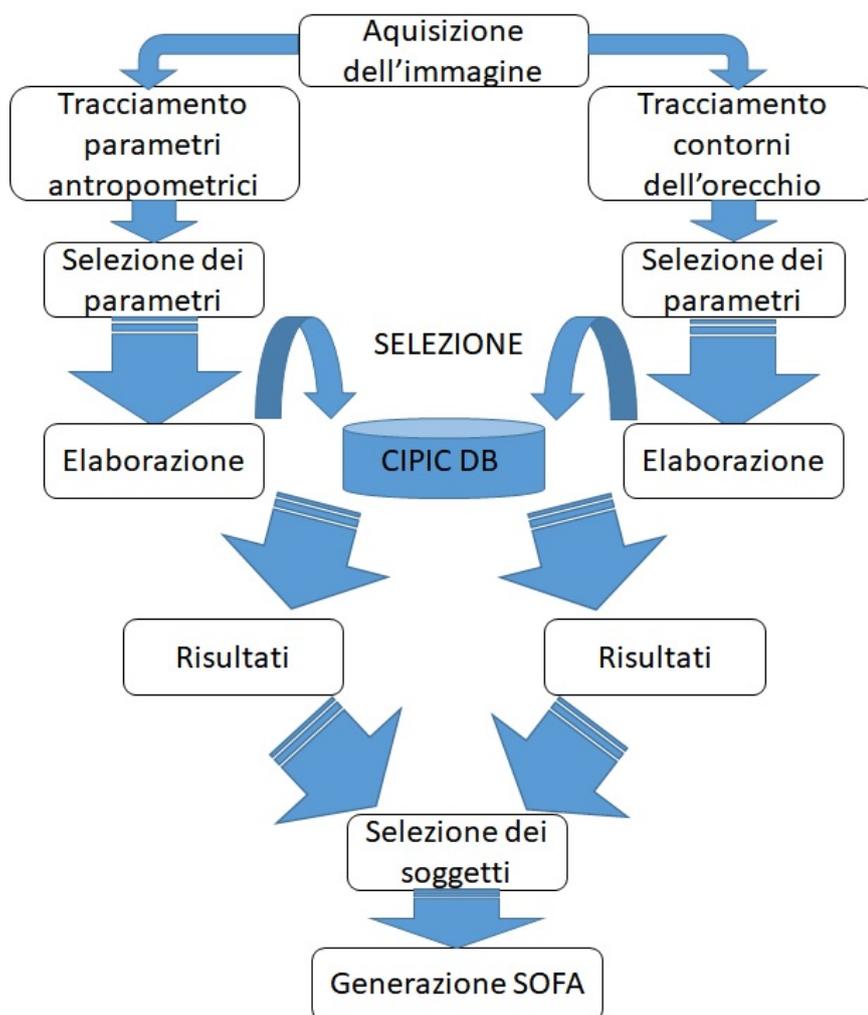


Figura 4.4: Schermata del menu' di generazione dei file Sofa. [10]

5

Test di localizzazione

Presentiamo in questo capitolo i confronti effettuati per validare il lavoro svolto in questa tesi, sia per supportare il lavoro precedentemente svolto, sia per avere delle conferme sulle migliorie apportate. Per eseguire questi test è stato creato un ambiente di realtà virtuale che permette la verifica della precisione nella localizzazione di sorgenti sonore. Quattordici volontari hanno preso parte a questo esperimento di localizzazione.

5.1 Descrizione del test di localizzazione

Il meccanismo di valutazione dei risultati, basato su alcuni lavori come [10][38], è un test che prevede la localizzazione di alcune fonti audio all'interno di un ambiente virtuale controllato, visualizzabile in Fig.5.1.

5.1.1 Hardware & software

Per questo test si è utilizzato un Desktop computer con le seguenti caratteristiche:

- OS: Windows 10 Pro
- Processore: Intel Core™ i7-3770K 3,50 GHz
- Ram: 8GB ddr3 1600MHz

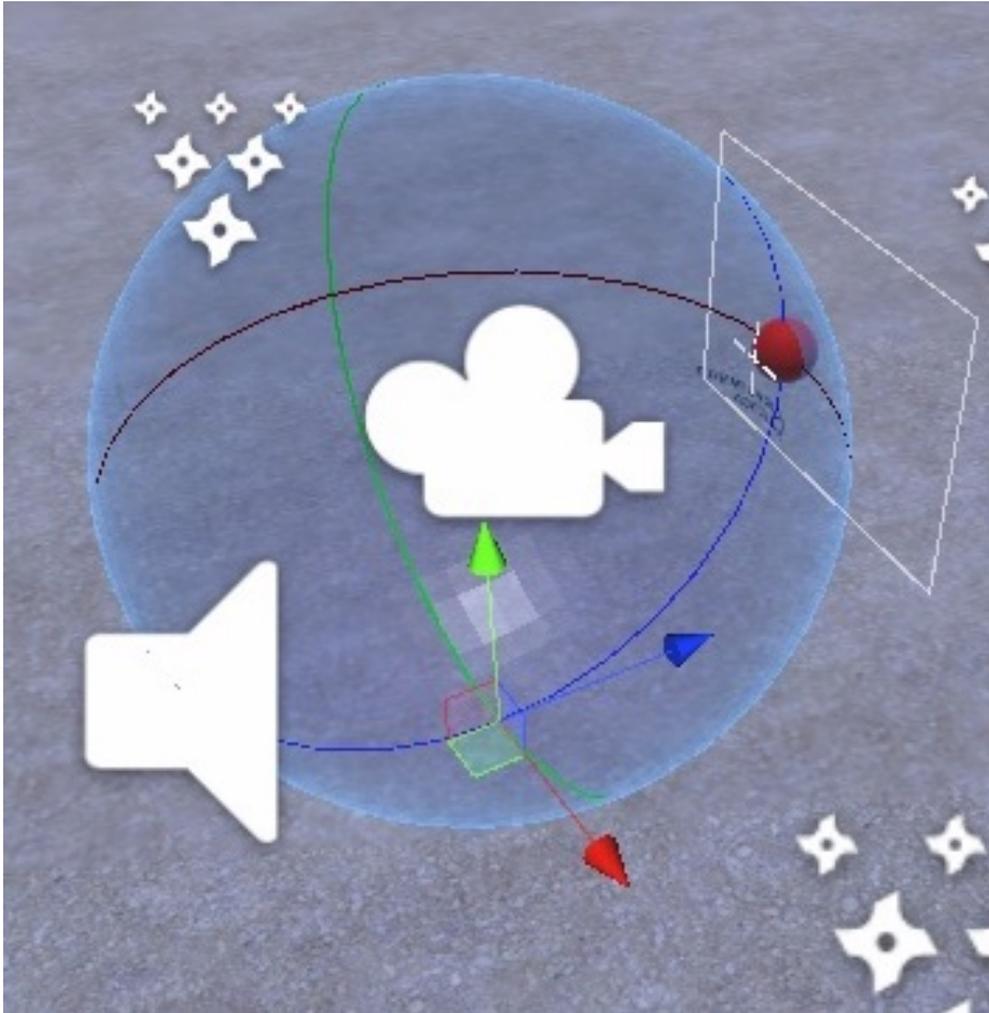


Figura 5.1: Schermata di Unity rappresentante lo scenario del test di localizzazione

- GPU: EVGA GeForce GTX 1060
- Hard-Disk: WD WD5000AAKX Blu 500GB

come elemento principale di elaborazione, e di un Samsung Galaxy S8 insieme al corrispettivo visore HMD Gear VR e controller per la realtà virtuale. Si è deciso di creare l'applicazione per la piattaforma HTC Vive, e poi emulare quest'ultima attraverso il programma VRidge¹. Questo ha permesso di avere a disposizione una potenza di calcolo maggiore e di far surriscaldare in modo minore il telefono, in più ha il vantaggio di elaborare l'audio direttamente dal computer, permettendo l'applicazione di

¹<https://riftcat.com/vridge>

filtri di equalizzazione per le cuffie in tempo reale. Le cuffie utilizzate sono state le Hefio, particolari cuffie in grado di auto-equalizzarsi attraverso una semplice procedura, e queste sono state collegate ad una scheda audio MOTU per l'output sonoro. L'applicazione è stata sviluppata utilizzando Unity 2018.3.² Per la spazializzazione del suono è stato usato Steam Audio 2.16³ visualizzabile in Fig.5.2, un renderer audio che permette la scelta di quale file di tipo SOFA si vuole utilizzare. La parte di realtà virtuale è stata gestita attraverso l'asset di SteamVR, un pacchetto per Unity che permette la completa gestione delle impostazioni per il 3D. Infine per la gestione dei livelli di equalizzazione e per la loro applicazione è stato usato EqualizerAPO⁴.

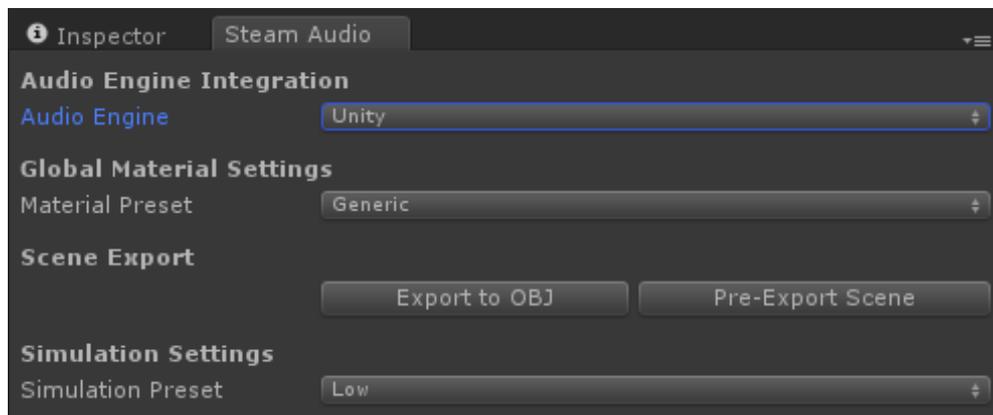


Figura 5.2: Schermata di Steam audio manager all'interno di Unity [11]

5.1.2 Ambiente

L'ambiente virtuale si presenta come un terreno di materiale sabbioso sotto ad un cielo azzurro dov'è stata situata una sfera di colore blu, nella quale sono stati evidenziati gli assi. L'utente si troverà al centro della sfera. Per segnalare l'origine del riferimento è stata usata una sfera di colore marrone. Per aiutare l'utente nella selezione delle posizione è stato utilizzato un mirino di colore bianco che punta sempre al centro del campo visivo. L'interfaccia grafica infine, da istruzione all'utente sulle azioni da intraprendere durante l'esecuzione.

²<https://unity3d.com/unity/whats-new/unity-2018.3.0>

³<https://valvesoftware.github.io/steam-audio/>

⁴<https://sourceforge.net/projects/equalizerapo/>

5.2 Motivazione dei confronti

I confronti portati si differenziano in base alle HRTF scelte e secondo la lunghezza dei test. Nei primi quattro test effettuati si è scelto di verificare che il numero di ripetizioni sia sufficiente e che dia risultati equivalenti ad uno di lunghezza maggiore, mentre nei rimanenti sono state testate le varie selezioni introdotte in questa tesi. Analizziamo in dettaglio quindi quali sono stati i confronti scelti in questo elaborato.

5.2.1 Metriche scelte

Per quanto riguarda la selezione basata sul padiglione dell'orecchio, basandoci sui risultati del lavoro precedente [16] [10], si è optato per la scelta della metrica con pesatura ottima con coefficiente di riflessione negativo, e un top-5 apparance.

Le selezioni per le misure antropometriche del viso introdotte sono in tutto 8, ma le prime tre si basano solo su una singola misura e non risultano molto preformanti, come non risulta molto valida la selezione basata sulla somma tra queste 3. Se si vuole quindi utilizzare una selezione basata puramente sul confronto tra metriche conviene quindi optare per l' A_{opt} , che attraverso la sua equazione permette una pesatura efficiente di tali valori. Nel caso di selezione per ITD la prima, quella basata sul Database, è stata scartata poiché non è stato possibile reperire informazioni sulla provenienza dei dati estratti direttamente dai file del CIPIC. La selezione tramite il modello di ITD genera dati analoghi a quelli generati dal solo A_{opt} , e si è quindi optato per la selezione antropometrica.

5.2.2 Confronto long e short

Per i primi dei due test si è deciso di eseguire un confronto tra i risultati di prove eseguiti per 73 ripetizioni, dette long, e quelle eseguite per 25 ripetizioni, dette short, per dimostrare che i risultati delle due prove sono simili. In questo primo confronto si è voluto verificare due ipotesi:

- numero di ripetizioni: si è voluto valutare se il numero di ripetizioni scelto avesse la stessa valenza, e quindi producesse risultati analoghi di un test molto più lungo, e se la ripetizione del test più volte producesse risultati migliori.

- precedente lavoro: si è portata una validazione del lavoro precedentemente svolto, replicandone le selezioni effettuate nella scelta delle HRTF [16] [10].

In questo modo è possibile effettuare test short in tutte le prove successive ottenendo gli stessi risultati dei long.

Per il secondo punto, si è deciso di utilizzare un'HRTF ricavata utilizzando il manichino Kemar, e di confrontarla con quella ricavata tramite selezione dei contorni dell'orecchio, come effettuato in [10], per confermare che questo tipo di selezione migliori la rilevazione dell'altezza.

Ricapitolando quindi attraverso i primi 4 confronti, andiamo a coprire i due punti sopraccitati, e per fare ciò, effettuiamo 3 prove long sia per l'HRTF kemar, che per l'HRTF selezionato con i contorni, e confrontiamo i risultati ottenuti con le medesime 2 prove eseguite con le 25 ripetizioni delle prove short.

5.2.3 A_{opt} e Cross-correlation

Questi due confronti sono basati sulle HRTF scelte attraverso le selezioni dell' A_{opt} e della cross-correlation. Nel primo caso si va quindi a cercare il soggetto del database le cui misure combinate danno il valore di A_{opt} più simile a quello calcolato per il nostro utente. Nel secondo si calcola tramite modello l'ITD della persona e si va a cercare il soggetto con i valori di ITD, estratti con cross-correlation, più simili a quello modellato. In questi due casi, la selezione di HRTF basata sull'ITD dovrebbe agire sulla percezione dell'azimuth, andandola a migliorare, mentre dovrebbe lasciare invariata quella dell'elevazione perché in questo caso nessun fattore per questa caratteristica è stato considerato. Come già specificato questo test è di tipo short.

5.2.4 Metodi combinati

In questi ultimi due casi, si vanno a combinare le due selezioni precedenti per ottenere tutti i vantaggi, e si va quindi a testare la funzionalità del plugin stesso. Si effettuano due test di tipo short sulle HRTF combinate ottenute attraverso lo script per la creazione di file SOFA. Si prende quindi il soggetto estratto con la selezione dei contorni come base, e si utilizzano gli ITD e gli ILD delle due selezioni testate prima per creare le HRTF per i confronti. In questo modo avremmo il vantaggio di migliorare sia la percezione di elevazione che quella di azimuth, ottenendo una localizzazione generale migliore.

5.3 Protocollo sperimentale

Descriviamo in questa sezione come si svolge L'esperimento, dividendo e spiegando i vari punti.

Acquisizione delle foto

Per prima cosa il soggetto viene fatto sedere su una sedia di posizione fissa calibrata per essere all'esatta distanza, nell'esatta direzione per scattare le foto e per avere sempre lo stesso riferimento, e viene poi aperta nel telefono, che successivamente viene posizionato nel visore, l'applicazione per l'orientamento. Questa permette attraverso gli assi di una sfera e un mirino di orientare la testa dell'utente in posizione perfettamente parallela e ortogonale alla camera. Dopo il posizionamento è possibile scattare le foto togliendo con attenzione il visore dal volto della persona, facendo attenzione di non spostarla. Questa procedura viene ripetuta sia per le foto frontali che laterali.

Elaborazione Matlab

Dopo aver inserito le foto nella cartella di Matlab, si procede con la creazione del profilo dell'utente. Per prima cosa si inseriscono i dati generali, e si va a concludere questa operazione con la selezione delle 3 foto.

Il passo successivo prevede di impostare il pixel-to-meter, per poi tracciare le misure antropometriche d_1 e d_5 dell'orecchio, le misure X_1 , X_2 e X_3 della testa e i contorni C_1 , C_2 e C_3 . A questo punto è possibile, dopo aver impostato i parametri, eseguire l'elaborazione dei risultati, disponibili nella cartella results. Per la selezione dei contorni si deve considerare l'Opt Weight Negative mentre per la testa si utilizzeranno le selezioni di model e di cross-correlation. Dopo aver ottenuto i risultati si utilizza il Generate SOFA per creare i file che saranno utilizzati successivamente.

Esperimento di localizzazione

Prima di cominciare il test vero e proprio, si consegna al soggetto il modulo contenente istruzioni e liberatoria. Dopo la visione e la firma di quest'ultimo si procede con la calibrazione delle cuffie Hefio, attraverso l'apposita applicazione per Android, e si inizia l'esperimento.

L'utente viene fatto accomodare all'interno della stanza silente per l'eliminazione dei rumori, come in Fig.5.3, dove indosserà il visore necessario per lo svolgimento del test. Subito dopo si troverà all'interno della sfera di colore blu, dove i movimenti possibili sono limitati a quelli della testa. Per iniziare l'ascolto è necessario mirare la sfera di colore marrone e premere il grilletto del controller, a quel punto verrà riprodotta una traccia audio composta da un treno di tre impulsi della durata totale di 3 secondi circa. Dopo l'ascolto il soggetto deve cercare di localizzare il punto d'origine della fonte sonora, e premere il grilletto del telecomando in dotazione per confermare la scelta. Fatto ciò, verrà salvato in un file la posizione del suono e la coordinata che l'utente ha selezionato, tutte e due espresse in valori di azimuth e di elevazione. Questo procedimento viene ripetuto per il numero di trial della prova.



Figura 5.3: Soggetto all'interno della camera silente pronto per effettuare il test.

Dato il notevole sforzo e la lunghezza dell'esperimento lungo, si è deciso di svolgere i test long e short in modo separato, a distanza di una settimana uno dall'altro. Alla fine della serie di 3 test long, e dei corrispettivi short si effettua un questionario per valutare il carico di lavoro, il NASA Task Load Index(TLX).

NASA Task Load Index

In questo questionario si valuta il carico di lavoro necessario al completamento dei test. Questo è diviso in due fasi, nella prima si valuta il carico di lavoro per ogni scala, con un punteggio da 0 a 20, nella seconda si definiscono i pesi delle scale tramite un confronto tra copie. Le sei scale di valutazioni sono accompagnate da una breve descrizione, e sono:

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

Finita la compilazione viene restituito un punteggio che stima una media totale dei risultati con valore massimo di 100 punti. Questo questionario serve per avere una stima del affaticamento dato dall'esperimento, soprattutto nel confronto tra i test con molte ripetizioni e quelli con poche.

6

Risultati

Illustriamo in questo capitolo i risultati del test di localizzazione effettuato su i seguenti 14 soggetti:

N.	ID Soggetto	Età	Data
1	122	57	10/09/2019
2	115	27	09/09/2019
3	125	24	10/09/2019
4	128	26	11/09/2019
5	123	26	10/09/2019
6	129	27	11/09/2019
7	127	27	11/09/2019
8	124	26	10/09/2019
9	118	20	09/09/2019
10	119	19	09/09/2019
11	120	27	09/09/2019
12	121	34	10/09/2019
13	126	27	10/09/2019
14	130	35	11/09/2019

Figura 6.1: Tabella dei soggetti partecipanti al test.

6.1 L'errore di localizzazione

Per il calcolo dei risultati ci siamo basati sulle differenze di rilevazione in modulo tra la posizione puntata e quella di origine, sia per i valori di azimuth che di elevazione, secondo la formula:

$$E_\theta = |\theta_r - \theta_u| \quad (6.1)$$

dove r e u indicano la posizione reale e quella dell'utente. Analogamente per l'elevazione abbiamo:

$$E_\phi = |\phi_r - \phi_u| \quad (6.2)$$

Nel caso della rilevazione dell'azimuth data la possibile presenza di errori di front-back, è stato deciso di correggere l'eq. 6.2 considerando un valore soglia di 90° , e di eliminarli secondo:

$$E_\theta = \begin{cases} 180 - E_\theta, & \text{if } E_\theta > 90 \\ E_\theta, & \text{if } E_\theta < 90 \end{cases} \quad (6.3)$$

Per finire è stata calcolata la media e la deviazione standard dei valori assoluti di errore per tutti i test di ogni soggetto e la media generale per ogni selezione calcolata a partire da questi dati.

6.2 Valutazione sull'errore medio

Valutiamo in questa sezione i test secondo quanto appena detto e riportiamo i risultati in tabella per trarne delle conclusioni.

6.2.1 Confronto Short-Long

Andiamo ad analizzare il primo dei confronti specificati nella sezione precedente, quello tra test di tipo long e short.

I risultati per l'errore medio sono disponibili in tabella 6.1, mentre la deviazione in 6.2.

N.	AVG. AZIMUTH				AVG. ELEVATION			
	KEMAR LONG	KEMAR SHORT	EAR LONG	EAR SHORT	KEMAR LONG	KEMAR SHORT	EAR LONG	CIPIC SHORT
1	11,3	11,32	12,12	14,9	26,38	27,55	23,57	23,32
2	14,37	15,24	13,87	12,91	28,9	27,57	24,75	23,92
3	20,59	19,47	17,46	17,46	28,27	26,58	28,11	27,54
4	33,16	32,57	33,08	33,05	28,97	29,65	29,45	30,29
5	16,87	18,04	18,84	18,14	31,7	30,34	27,29	28,27
6	20,42	19,2	20,46	21,59	30,7	30,68	32,72	31,38
7	12,26	12,5	11,84	11,87	30,01	29,63	24,52	24,92
8	21,38	21,69	20,98	20,45	26,98	27,67	24,9	25,19
9	20,27	21,19	18,53	18,11	26,83	26,92	22,8	23,53
10	29,22	29,5	22,84	23,74	28,18	27,23	27,32	24,93
11	12,64	12,03	12,74	11,21	28,68	26,12	25,73	24,63
12	31,13	30,58	25,02	24,27	29,11	28,53	27,22	26,89
13	13,44	12,72	12,04	12,66	27,57	25,4	26,26	23,81
14	14,32	11,83	12,39	14,87	23,13	24,45	22,65	22,51
AVG.	19,38	19,13	18,02	18,23	28,24	27,74	26,24	25,80

Tabella 6.1: Errore medio assoluto di azimuth e di elevazione tra kemar e cipic sia per i test long che short.

N.	ST.DEV. AZIMUTH				ST.DEV. ELEVATION			
	KEMAR LONG	KEMAR SHORT	EAR LONG	EAR SHORT	KEMAR LONG	KEMAR SHORT	CIPIC LONG	CIPIC SHORT
1	14,55	12,23	11,51	18,52	21,72	19,73	20,88	18,93
2	14,72	13,07	13,61	13,86	22,85	17,16	19,94	13,64
3	18,3	20,28	15,86	21,29	23,45	21,46	24,88	19,73
4	23,43	21,75	23,69	23,77	22,03	17,58	21,12	22,51
5	16,94	16,47	17,86	15,73	26,27	22,19	21,33	21,49
6	19,53	14,07	20,16	20,8	22,29	19,01	25,26	28,21
7	11,07	12,75	9,38	11,59	24,21	20,12	21,86	15,44
8	21,82	16,11	20,33	13,88	21,5	21,58	22,48	22,79
9	18,62	20,68	19,56	14,09	19,82	18,79	15,28	16,34
10	20,5	25,62	20,23	20,1	18,81	19,97	17,83	15,86
11	15	10,54	14,6	14,88	23,27	20,87	19,86	22,09
12	26,75	21,64	23,6	20,59	20,8	15,95	20,23	23,55
13	14,18	12,59	12,23	15,18	18,87	19,16	17,54	15,08
14	11,8	12,06	12,34	19,63	19,63	18,36	19,12	19,23
AVG.	17,66	16,42	16,76	16,90	21,82	19,42	20,54	19,64

Tabella 6.2: Deviazione standard di azimuth e di elevazione tra kemar e cipic sia per i test long che short.

Analizzando i risultati è evidente che i due fattori chiave che ci aspettavamo, cioè la conferma dei risultati del precedente lavoro e il fatto che test di una maggiore durata producano risultati paragonabili a prove più brevi, sono stati verificati. Per il primo caso, otteniamo un miglioramento di due gradi nella localizzazione verticale dei suoni mentre non ricaviamo benefici per quanto riguarda la localizzazione orizzontale in linea con [10]. Per il secondo caso, la sessione di 3 test long ha un errore medio molto simile a quello del singolo esperimento di tipo short. Questo ci permette di procedere con il confronto tra metodi di selezione utilizzando solo le prove di tipo breve, in accordo con l'ipotesi iniziale.

Curva di apprendimento

Possiamo servirci dei test long per controllare se c'è stato un miglioramento dovuto alle serie di prove ripetute. Per fare questo è possibile analizzare la media d'errore per le singole prove lunghe. I risultati sono riportati in tabella 6.3.

Essendo i risultati molto simili tra loro si può concludere che non sono presenti miglioramenti dovuti all'esperienza maturata nelle prove. Va anche aggiunto che la lunghezza di una sessione intera di 3 test long è equivalente a quella di tutti gli

	Kemar			EAR SEL.				Kemar			EAR SEL.		
	1	2	3	1	2	3		1	2	3	1	2	3
AVG. AZ	19,65	19,21	19,34	17,94	18,42	18,05	ST.DEV. AZ	7,26	7,19	6,68	5,74	7,12	5,48
AVG. EL	28,32	28,69	27,72	27,16	26,28	27,31	ST.DEV. EL	1,82	2,39	2,62	2,34	3,67	2,78

Tabella 6.3: Risultati delle 3 ripetizioni per Kemar e la selezione dei contorni.

short, per cui possiamo dire che anche in questo caso non abbiamo condizionamenti introdotti da un'eventuale curva di apprendimento.

6.2.2 Confronti tra selezioni

Passiamo ora ai confronti per validare il principale lavoro di questa tesi. Nelle seguenti tabelle sono presenti tutti i risultati sulle varie selezione e composizioni di HRTF. La 6.4 mostra i risultati ottenuti confrontando l'errore in azimuth mentre la 6.4 mostra la deviazione standard . In questo caso si nota che le selezioni basate sulla personalizzazione dell'ITD migliorano la localizzazione orizzontale di un valore significativo, e questa miglioramento persiste anche nelle HRTF create tramite il tool elaborato in questo lavoro. Questo dato può essere visualizzato anche nel grafico 6.2 Per quanto riguarda l'elevazione i risultati sono riportati in 6.5 e in 6.8. Qui è la tecnica di selezione dei contorni a produrre risultati visibili, e questo influisce sia sulla HRTF direttamente selezionata con questo metodo sia sulle HRTF combinate con le due selezioni. L'istogramma, presente in figura 6.3 conferma i risultati tra il Kemar e la selezione dei contorni, ma lo stesso non vale per i metodi di selezione dell'ITD, dove il non si vede una netta distinzione tra i metodi di selezione. Il metodo di selezione combinato invece sembra fare meglio anche in questo caso.

AVG. AZIMUTH						
N.	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	11,32	14,9	10,52	11,04	9,51	7,47
2	15,24	12,91	10,89	7,79	8,09	8,93
3	19,47	17,46	13,51	14,21	13,14	12,03
4	32,57	33,05	25,48	31,55	32,55	23,93
5	18,04	18,14	16,15	12,97	12,75	15,31
6	19,2	21,59	15,7	17,7	17,14	14,15
7	12,5	11,87	13,48	12,52	11,18	10,19
8	21,69	20,45	18,78	19,23	12,3	19,87
9	21,19	18,11	15,97	19,16	23,72	19,6
10	29,5	23,74	16,95	20,66	20,45	18,17
11	12,03	11,21	11,32	7,67	8,6	8,63
12	30,58	24,27	22,08	22,71	20,95	21,15
13	12,72	12,66	9,83	9,54	8,14	10,43
14	11,83	14,87	11,05	8,81	10,02	11,53
AVG.	19,13	18,23	15,12	15,40	14,90	14,39

Tabella 6.4: Errore medio assoluto nella rilevazione del azimuth per le varie selezioni.

AVG. ELEVATION						
N.	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	27,55	23,32	28,4	25,65	23,13	23,01
2	27,57	23,92	27,7	31,07	26,7	22,33
3	26,58	27,54	26,37	28,39	25,1	26,12
4	29,65	30,29	31,12	34,55	30,16	29,9
5	30,34	28,27	26,1	31,9	27,88	30,47
6	30,68	31,38	32,03	29,2	28,74	30,13
7	29,63	24,92	26,21	23	24,62	22,74
8	27,67	25,19	26,81	29,35	22,44	23,52
9	26,92	23,53	23,16	23,38	23,02	21,83
10	27,23	24,93	23,89	19,31	22,34	18,1
11	26,12	24,63	23,09	30,63	20,96	18,25
12	28,53	26,89	26,28	26,15	27,13	25,03
13	25,4	23,81	23,15	24,47	20,93	21,72
14	24,45	22,51	26,25	21,26	21,95	23,72
AVG.	27,74	25,80	26,47	27,02	24,65	24,06

Tabella 6.5: Errore medio assoluto nella rilevazione dell'elevazione per le varie selezioni.

Std. AZIMUTH						
N.	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	12,23	18,52	8,74	17,83	16,41	7,53
2	13,07	13,86	13,38	10,04	8,01	8,68
3	20,28	21,29	14,68	19,22	19,22	9,06
4	21,75	23,77	20,59	25,08	28,01	25,53
5	16,47	15,73	20,99	16,23	17,87	21,8
6	14,07	20,8	18,86	17,84	16,88	14,44
7	12,75	11,59	15,54	11,31	13,57	14,67
8	16,11	13,88	16,27	13,68	11,72	19,5
9	20,68	14,09	20,52	16,09	15,89	22,15
10	25,62	20,1	13,92	20,13	20	15,49
11	10,54	14,88	12,96	9,9	10,5	12,38
12	21,64	20,59	20,97	19,58	19,07	23,06
13	12,59	15,18	9,01	13,03	12,17	18,98
14	12,06	12,34	11,87	10,77	12,11	11,64
AVG.	16,42	16,90	15,59	15,77	15,82	16,07

Tabella 6.6: Deviazione standard nella rilevazione del azimuth per le varie selezioni.

N.	Std. ELEVATION					
	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	19,73	18,93	21,97	20,6	22,49	23,03
2	17,16	13,64	24,65	21,84	15,69	14,55
3	21,46	19,73	20,11	24,35	22,49	22,28
4	17,58	22,51	25,68	28,46	25,35	24,63
5	22,19	21,49	25,1	19,52	26,67	22,88
6	19,01	28,21	23,68	20,8	21,51	20,56
7	20,12	15,44	20,1	19,4	16,46	16,22
8	21,58	22,79	14,77	26,16	18,84	23,19
9	18,79	16,34	18,33	15,05	15,45	16,83
10	19,97	15,86	16,26	18,05	15,57	18,7
11	20,87	22,09	22,42	22,58	16,8	14,27
12	15,95	23,55	20,51	22,74	22,93	22,29
13	19,16	15,08	15,05	14,08	18,99	19,32
14	18,36	19,23	20,66	17,57	19,57	24,73
AVG.	19,42	19,64	20,66	20,80	19,92	20,25

Tabella 6.7: Deviazione standard nella rilevazione dell'elevazione per le varie selezioni.

N.	Std. ELEVATION					
	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	19,73	18,93	21,97	20,6	22,49	23,03
2	17,16	13,64	24,65	21,84	15,69	14,55
3	21,46	19,73	20,11	24,35	22,49	22,28
4	17,58	22,51	25,68	28,46	25,35	24,63
5	22,19	21,49	25,1	19,52	26,67	22,88
6	19,01	28,21	23,68	20,8	21,51	20,56
7	20,12	15,44	20,1	19,4	16,46	16,22
8	21,58	22,79	14,77	26,16	18,84	23,19
9	18,79	16,34	18,33	15,05	15,45	16,83
10	19,97	15,86	16,26	18,05	15,57	18,7
11	20,87	22,09	22,42	22,58	16,8	14,27
12	15,95	23,55	20,51	22,74	22,93	22,29
13	19,16	15,08	15,05	14,08	18,99	19,32
14	18,36	19,23	20,66	17,57	19,57	24,73
AVG.	19,42	19,64	20,66	20,80	19,92	20,25

Tabella 6.8: Deviazione standard nella rilevazione dell'elevazione per le varie selezioni.

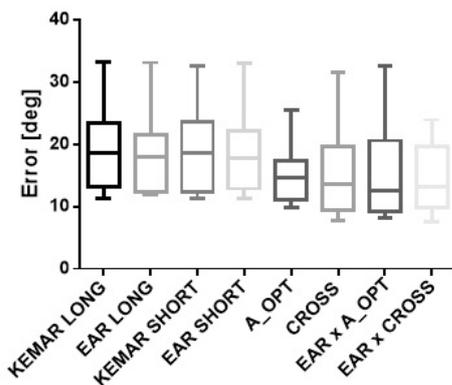


Figura 6.2: Istogramma dei dati di errore medio assoluto per l'azimuth.

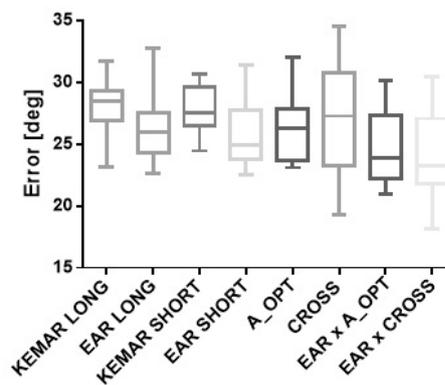


Figura 6.3: Istogramma dei dati di errore medio assoluto per l'elevazione.

6.2.3 Errori di Front-Back

Analizzando gli errori di front-back, considerando per i long la media delle 3 prove, otteniamo la tabella 6.9. Qui otteniamo che le medie tra long e short, essendo i long 3 volte più lunghi di quest'ultimi sono perfettamente bilanciate. Si nota inoltre che i soggetti 4 e 12 hanno effettuato un numero di errori 3 volte superiore alla media.

N.	FRONT-BACK							
	KEMAR LONG	CIPIC LONG	KEMAR SHORT	CIPIC SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	3,0	1,3	0	0	0	2	1	1
2	2,3	1,3	1	0	0	0	1	0
3	2,3	2,0	0	1	0	2	1	1
4	9,7	7,0	3	4	3	3	1	2
5	2,7	1,7	2	1	2	2	2	2
6	3,7	3,3	0	0	2	0	0	2
7	1,3	1,0	0	0	1	0	1	1
8	3,3	4,0	0	0	1	0	0	2
9	3,3	4,0	0	0	1	0	2	2
10	4,7	4,7	2	1	1	1	1	1
11	3,3	2,7	2	4	1	1	2	2
12	9,3	8,3	1	0	0	3	2	2
13	1,7	1,3	1	0	0	1	1	1
14	2,7	1,7	0	0	1	1	0	1
AVG.	3,81	3,17	0,86	0,79	0,93	1,14	1,07	1,43

Tabella 6.9: Numero di front-back per ogni condizione e per ogni soggetto.

6.2.4 Eliminazione delle anomalie

Per ottenere dati più omogenei, in linea con il precedente lavoro [10], procediamo con l'eliminazione di 3 soggetti che hanno effettuato troppi errori di front-back o per i quali la differenza dei valori tra alcune selezioni risultava troppo elevato. Eliminando dunque i soggetti numero 4,10 e 12 otteniamo tabelle le 6.10 6.11 6.12 6.13

N.	AVG. AZIMUTH							
	KEMAR LONG	EAR LONG	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	11,3	12,12	11,32	14,9	10,52	11,04	9,51	7,47
2	14,37	13,87	15,24	12,91	10,89	7,79	8,09	8,93
3	20,59	17,46	19,47	17,46	13,51	14,21	13,14	12,03
4	31,16	31,08	32,57	31,05	35,38	31,55	32,55	31,91
5	16,87	18,84	18,04	18,14	16,15	12,97	12,75	15,31
6	20,42	20,46	19,2	21,59	15,7	17,7	17,14	14,15
7	12,26	11,84	12,5	11,87	13,48	12,52	11,18	10,19
8	21,38	20,98	21,69	20,45	18,78	19,23	12,3	19,87
9	20,27	18,53	21,19	18,11	15,97	19,16	23,72	19,6
10	39,32	37,89	39,5	31,78	36,95	30,66	30,45	38,17
11	12,64	12,74	12,03	11,21	11,32	7,67	8,6	8,63
12	31,13	25,02	30,58	24,27	22,08	22,71	20,95	21,15
13	13,44	12,04	12,72	12,66	9,83	9,54	8,14	10,43
14	14,32	12,39	11,83	14,87	11,05	8,81	10,02	11,53
AVG.	16,17	15,57	15,93	15,83	13,38	12,79	12,24	12,56

Tabella 6.10: Errore medio assoluto nella rilevazione del azimuth per le varie selezioni eliminando i soggetti 4,10 e 12.

N.	AVG. ELEVATION							
	KEMAR LONG	EAR LONG	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	26,38	23,57	27,55	23,32	28,4	25,65	23,13	23,01
2	28,9	24,75	27,57	23,92	27,7	31,07	26,7	22,33
3	28,27	28,11	26,58	27,54	26,37	28,39	25,1	26,12
4	26,97	29,65	28,65	26,75	31,17	26,75	29,15	28,4
5	31,7	27,29	30,34	28,27	26,1	31,9	27,88	30,47
6	30,7	32,72	30,68	31,38	32,03	29,2	28,74	30,13
7	30,01	24,52	29,63	24,92	26,21	23	24,62	22,74
8	26,98	24,9	27,67	25,19	26,81	29,35	22,44	23,52
9	26,83	22,8	26,92	23,53	23,16	23,38	23,02	21,83
10	26,15	27,27	27,23	25,75	23,85	19,71	22,29	18,1
11	28,68	25,73	26,12	24,63	23,09	30,63	20,96	18,25
12	27,51	26,8	26,8	23,89	26,28	28,15	27,15	25,25
13	27,57	26,26	25,4	23,81	23,15	24,47	20,93	21,72
14	23,13	22,65	24,45	22,51	26,25	21,26	21,95	23,72
AVG.	28,10	25,75	27,54	25,37	26,30	27,12	24,13	23,99

Tabella 6.11: Errore medio assoluto nella rilevazione dell'elevazione per le varie selezioni eliminando i soggetti 4,10 e 12.

N.	Std. AZIMUTH							
	KEMAR LONG	EAR LONG	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	14,55	11,51	12,23	18,52	8,74	17,83	16,41	7,53
2	14,72	13,61	13,07	13,86	13,38	10,04	8,01	8,68
3	18,3	15,86	20,28	21,29	14,68	19,22	19,22	9,06
4	17,44	17,87	17,7	19,77	17,88	19,88	19,07	21,8
5	16,94	17,86	16,47	15,73	20,99	16,23	17,87	21,8
6	19,53	20,16	14,07	20,8	18,86	17,84	16,88	14,44
7	11,07	9,38	12,75	11,59	15,54	11,31	13,57	14,67
8	21,82	20,33	16,11	13,88	16,27	13,68	11,72	19,5
9	18,62	19,56	20,68	14,09	20,52	16,09	15,89	22,15
10	19,5	19,77	20,67	19,1	13,84	20,13	20	15,49
11	15	14,6	10,54	14,88	12,96	9,9	10,5	12,38
12	16,75	19,6	21,64	20,59	20,97	19,58	19,07	23,06
13	14,18	12,23	12,59	15,18	9,01	13,03	12,17	18,98
14	11,8	12,06	12,06	12,34	11,87	10,77	12,11	11,64
AVG.	16,05	15,20	14,62	15,65	14,80	14,18	14,03	14,62

Tabella 6.12: Deviazione standard nella rilevazione del azimuth per le varie selezioni eliminando i soggetti 4,10 e 12.

N.	Std. ELEVATION							
	KEMAR LONG	EAR LONG	KEMAR SHORT	EAR SHORT	A_OPT	CROSS	EAR x A_OPT	EAR x CROSS
1	21,72	20,88	19,73	18,93	21,97	20,6	22,49	23,03
2	22,85	19,94	17,16	13,64	24,65	21,84	15,69	14,55
3	23,45	24,88	21,46	19,73	20,11	24,35	22,49	22,28
4	22,09	21,17	17,98	22,51	20,88	28,86	20,35	24,64
5	26,27	21,33	22,19	21,49	25,1	19,52	26,67	22,88
6	22,29	25,26	19,01	28,21	23,68	20,8	21,51	20,56
7	24,21	21,86	20,12	15,44	20,1	19,4	16,46	16,22
8	21,5	22,48	21,58	22,79	14,77	26,16	18,84	23,19
9	19,82	15,28	18,79	16,34	18,33	15,05	15,45	16,83
10	19,85	17,87	19,87	19,85	19,26	19,85	19,7	18,7
11	23,27	19,86	20,87	22,09	22,42	22,58	16,8	14,27
12	20,8	20,23	19,95	19,55	20,51	12,74	22,03	23,25
13	18,87	17,54	19,16	15,08	15,05	14,08	18,99	19,32
14	19,63	19,12	18,36	19,23	20,66	17,57	19,57	24,73
AVG.	22,17	20,77	19,86	19,36	20,62	20,18	19,54	19,81

Tabella 6.13: Deviazione standard nella rilevazione dell'elevazione per le varie selezioni eliminando i soggetti 4,10 e 12.

6.3 Risultati Nasa TLX

Riportiamo qui i grafici contenenti i dati ricavati dai questionari Nasa TLX per ogni soggetto. È facile verificare che le scale con valore maggiore sono quella della richiesta mentale e quella delle performance, seguite da quella di richiesta temporale. Si può

anche concludere che il punteggio finale sul carico di lavoro sia nella maggior parte dei casi maggiore per i test di tipo long piuttosto che in quelli di tipo short, come mostrato dalla Tab. 6.14 e dal grafico 6.18

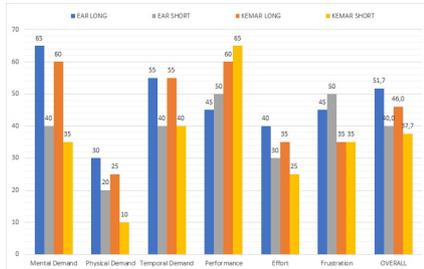


Figura 6.4: NasaTLX Soggetto 1

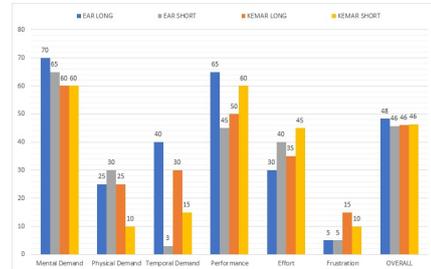


Figura 6.5: NasaTLX Soggetto 2

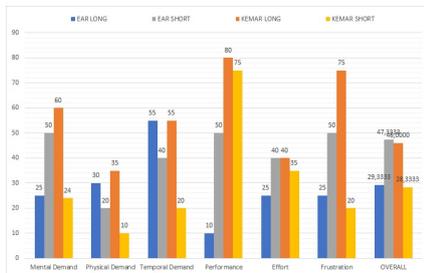


Figura 6.6: NasaTLX Soggetto 3

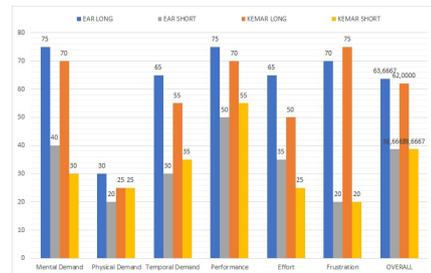


Figura 6.7: NasaTLX Soggetto 4

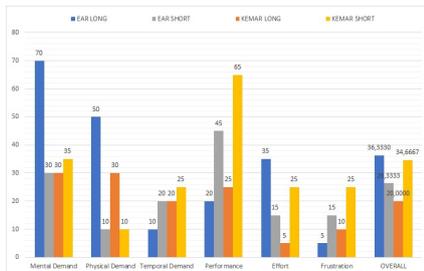


Figura 6.8: NasaTLX Soggetto 5

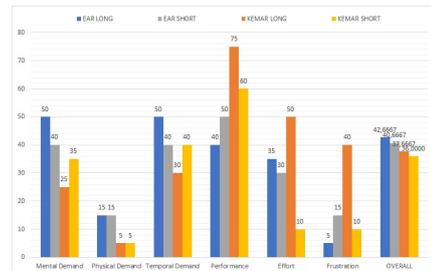


Figura 6.9: NasaTLX Soggetto 6

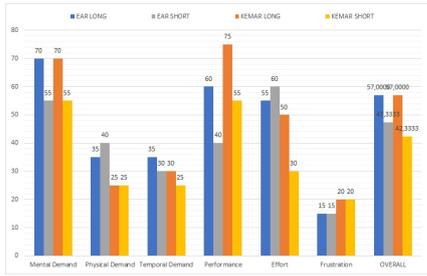


Figura 6.10: NasaTLX Sogetto 7

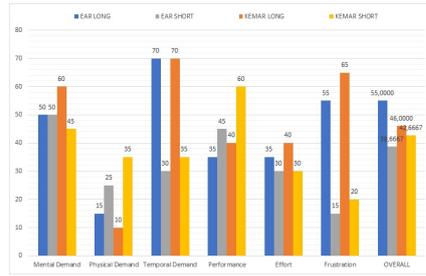


Figura 6.11: NasaTLX Sogetto 8

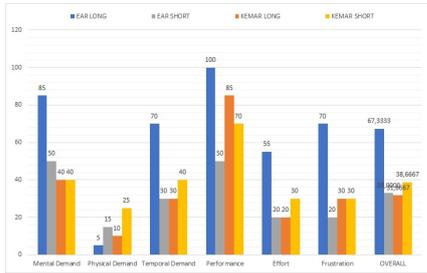


Figura 6.12: NasaTLX Sogetto 9

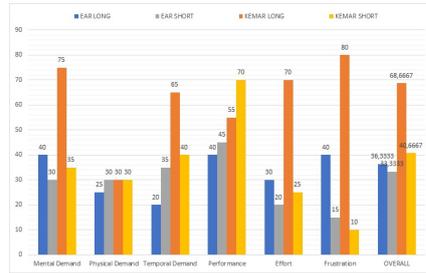


Figura 6.13: NasaTLX Sogetto 10

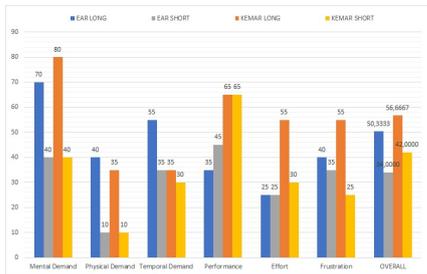


Figura 6.14: NasaTLX Sogetto 11

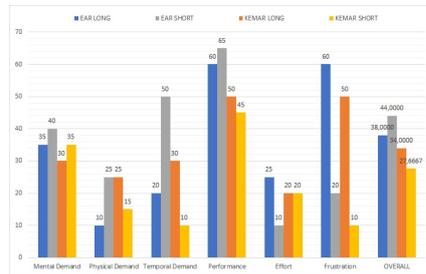


Figura 6.15: NasaTLX Sogetto 12

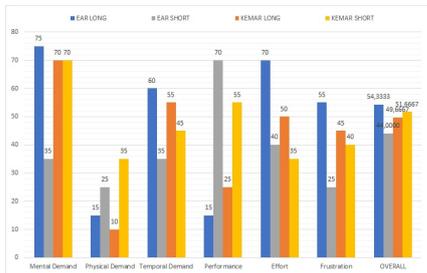


Figura 6.16: NasaTLX Sogetto 13

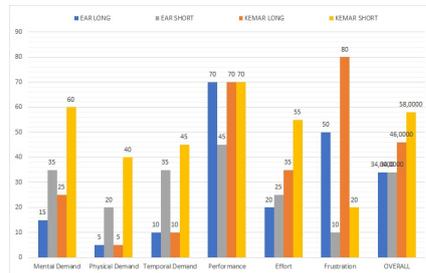


Figura 6.17: NasaTLX Sogetto 14

	AVG. NASATLX				ST.DEV. NASATLX			
	EAR LONG	EAR SHORT	KEMAR LONG	KEMAR SHORT	EAR LONG	EAR SHORT	KEMAR LONG	KEMAR SHORT
Mental Demand	56,79	42,86	53,93	42,79	56,79	42,86	53,93	42,79
Physical Demand	23,57	23,21	21,07	21,79	23,57	23,21	21,07	21,79
Temporal Demand	43,93	32,36	40,71	31,79	43,93	32,36	40,71	31,79
Performance	47,86	49,64	58,93	62,14	47,86	49,64	58,93	62,14
Effort	41,79	30,00	39,64	30,00	41,79	30,00	39,64	30,00
Frustration	38,57	22,14	48,21	21,07	38,57	22,14	48,21	21,07
OVERALL	48,88	39,07	46,24	40,38	48,88	39,07	46,24	40,38

Tabella 6.14: Tabella per la media e varianza dei risultati dei Nasa TLX

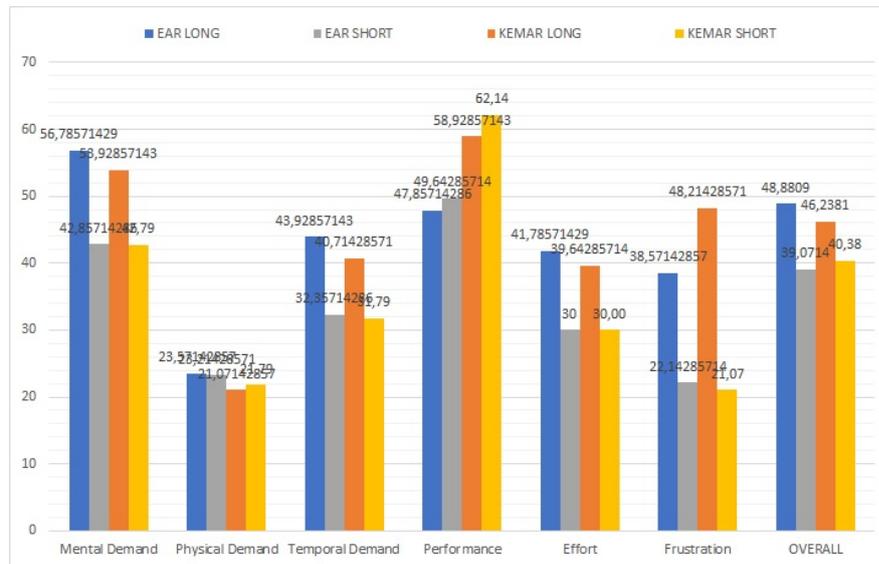


Figura 6.18: Grafico per la media dei risultati dei Nasa TLX

7

Conclusioni

In questa tesi ci siamo posti lo scopo di migliorare la selezione di HRTF basandoci sulle foto dei soggetti. Per fare ciò, si è partiti dal lavoro precedente sulla selezione dei contorni dell'orecchio che è ampliato includendo le selezioni tramite misure antropometriche, somiglianza di ITD e la possibilità di scegliere quale ITD utilizzare nell'HRTF risultante.

Andando quindi ad analizzare i risultati si è visto come esperimenti molto lunghi generino risultati analoghi a quelli più corti, com'è stato dimostrato nei confronti tra long e short, ma quest'ultimi sono meno onerosi da sostenere sia in termini temporali, sia in termini di carico di lavoro come si può vedere dai risultati dei Nasa TLX. Il confronto tra HRTF invece ha ribadito che la selezione del precedente lavoro migliora la localizzazione sul piano verticale rispetto al Kemar, ma rimangono ancora dei dubbi, come visto dai risultati della figura 6.3, sulle misure legate alle selezioni dell'ITD. Per verificare questo aspetto sarebbe necessario ripetere i test con più soggetti, e questo potrebbe essere l'obiettivo di futuri lavori su questa tesi, mentre si può concludere che quelle basate sulla selezione del volto migliorano quella del piano orizzontale. Le due selezioni prese in esame hanno generato valori di errore molto simili tra loro, questo è probabilmente dovuto alla piccola variazione che si ha tra i due ITD selezionati. Infine possiamo concludere che andando ad unire in un'unica HRTF le selezioni dei contorni con una per l'ITD è stato possibile ottenere un miglioramento della localizzazione generale del suono sia nel piano verticale che orizzontale, riuscendo a ridurre

gli errori in azimuth e in elevazione con tutte e due le HRTF create.

Possiamo quindi dire che questo lavoro può essere una buona base di partenza per lo sviluppo di software per la selezione di HRTF personalizzate. Si potrebbe rendere ancora più user-friendly la selezione di parametri o il calcolo delle misure per estendere un ipotetico software a utenti senza alcuna conoscenza sulle tematiche descritte in questa tesi, andando a generare file da implementare direttamente nelle applicazioni o nei videogiochi per avere ambienti di realtà virtuale più realistici e immersivi oppure per la maggior precisione nella localizzazione in videogiochi più frenetici e veloci.

Bibliografia

- [1] J. Eliasson, U. Leijon, and E. Persson, “Multichannel cinema sound,” *Lulea Tekniska Universitet, SMD074, Distributed multimedia*, 2001.
- [2] S. Mattes, P. Nelson, F. Fazi, and M. Capp, “Towards a human perceptual model for 3d sound localization,” vol. 34, 11 2012.
- [3] F. S. Woods, *Higher geometry: an introduction to advanced methods in analytic geometry*. Courier Corporation, 2013.
- [4] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [5] M. Geronazzo, J. Kleimola, and P. Majdak, “Personalization support for binaural headphone reproduction in web browsers,” 01 2015.
- [6] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The cipic hrtf database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, Oct 2001, pp. 99–102.
- [7] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, “Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4463–4467.
- [8] H. Bahu and D. Romblo, “Optimization and prediction of the spherical and ellipsoidal itd model parameters using offset ears,” 07 2018.
- [9] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model,” *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

- [10] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, “Improving elevation perception with a tool for image-guided head-related transfer function selection,” in *Proc. of the 20th Int. Conference on Digital Audio Effects (DAFx-17)*, 2017, pp. 397–404.
- [11] Steam audio. [Online]. Available: <https://valvesoftware.github.io/steam-audio>
- [12] A. Andreopoulou and B. F. Katz, “Identification of perceptually relevant methods of inter-aural time difference estimation,” *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 588–598, 2017.
- [13] Vola il consumo di videogiochi (+18%) trainato dalle vendite digitali. [Online]. Available: <https://www.ilsole24ore.com/art/vola-consumo-videogiochi-18percento-trainato-vendite-digitali---ABAQ8jiB>
- [14] Introduction to virtual reality audio. [Online]. Available: <https://developer.oculus.com/documentation/audiosdk/latest/concepts/book-audio-intro>
- [15] E. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” *The Journal of the Acoustical Society of America*, vol. 94, pp. 111–23, 08 1993.
- [16] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, “Applying a single-notch metric to image-guided head-related transfer function selection for improved vertical localization,” *J. Audio Eng. Soc*, vol. 67, no. 6, pp. 414–428, 2019. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=20483>
- [17] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Springer Science & Business Media, 2013, vol. 22.
- [18] C. Cheng and G. Wakefield, “Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space,” *AES: Journal of the Audio Engineering Society*, vol. 49, pp. 231–249, 04 2001.
- [19] M. Geronazzo, S. Spagnol, and F. Avanzini, “Mixed structural modeling of head-related transfer functions for customized binaural audio delivery,” in *2013 18th International Conference on Digital Signal Processing (DSP)*. IEEE, 2013, pp. 1–8.

- [20] Y. Iwaya, “Individualization of head-related transfer functions with tournament-style listening test: Listening with other’s ears,” *Acoustical Science and Technology*, vol. 27, no. 6, pp. 340–343, 2006.
- [21] L. Sarlat, O. Warusfel, and I. Viaud-Delmon, “Ventriloquism aftereffects occur in the rear hemisphere,” *Neuroscience letters*, vol. 404, pp. 324–9, 10 2006.
- [22] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, “Hrtf personalization using anthropometric measurements,” in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, Oct 2003, pp. 157–160.
- [23] G. F. Kuhn, “Model for the interaural time differences in the azimuthal plane,” *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977. [Online]. Available: <https://doi.org/10.1121/1.381498>
- [24] E. Macpherson and J. Middlebrooks, “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,” *The Journal of the Acoustical Society of America*, vol. 111, pp. 2219–36, 06 2002.
- [25] L. Rayleigh, “Xii. on our perception of sound direction,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.
- [26] R. S. Woodworth and H. Schlosberg, “Experimental psychology, rev,” 1954.
- [27] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, “Approximating the head-related transfer function using simple geometric models of the head and torso,” *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002. [Online]. Available: <https://doi.org/10.1121/1.1508780>
- [28] J. Middlebrooks, “Individual differences in external-ear transfer functions reduced by scaling in frequency. journal of the acoustical society of america 106:1480-1492,” *The Journal of the Acoustical Society of America*, vol. 106, pp. 1480–92, 10 1999.

- [29] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, “The interaural time difference in binaural synthesis,” in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.
- [30] M. Geronazzo, J. Kleimola, and P. Majdak, “Personalization support for binaural headphone reproduction in web browsers,” 01 2015.
- [31] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, “Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions,” 05 2013.
- [32] P. Satarzadeh, “A study of physical and circuit models of the human pinnae,” Ph.D. dissertation, University of California, Davis, 2006.
- [33] R. Lord, “The theory of sound,” 1945.
- [34] C. P. Brown and R. O. Duda, “A structural model for binaural sound synthesis,” *IEEE transactions on speech and audio processing*, vol. 6, no. 5, pp. 476–488, 1998.
- [35] V. R. Algazi, C. Avendano, and R. O. Duda, “Estimation of a spherical-head model from anthropometry,” *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–479, 2001.
- [36] R. Bomhardt, *Anthropometric Individualization of Head-Related Transfer Functions Analysis and Modeling*. Logos Verlag Berlin GmbH, 2017, vol. 28.
- [37] R. O. Duda, V. R. Algazi, and D. M. Thompson, “The use of head-and-torso models for improved spatial sound synthesis,” in *Audio Engineering Society Convention 113*. Audio Engineering Society, 2002.
- [38] M. Geronazzo, E. Sikström, J. Kleimola, F. Avanzini, A. De Götzen, and S. Serafin, “The impact of an accurate vertical localization with hrtfs on short explorations of immersive virtual reality scenarios,” in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2018, pp. 90–97.