



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
Corso di laurea triennale in Ingegneria Elettronica

Misura ed Analisi delle Head Related Transfer Functions in Ambienti Reali per Sistemi Audio Immersivi

Measurement and Analysis of Head Related Transfer Functions in Real Environments for Immersive Audio Systems

Tesi di Laurea di:
Francisco Messina

Relatore: Chiar.mo
Prof. Stefania Cecchi

Correlatore:
Ing. Alessandro Terenzi

Anno Accademico 2021/2022

Abstract

I sistemi di riproduzione audio, al giorno d'oggi, non riescono a fornire una rappresentazione tridimensionale del suono. Grazie all'analisi e comprensione delle Head Related Transfer Function (HRTF), è possibile implementare algoritmi sempre più efficienti per creare ambienti acustici virtuali tridimensionali, utilizzando cuffie o altoparlanti.

Lo scopo di questa tesi è quello di effettuare uno studio sulla risposta del sistema uditivo umano in relazione ai fenomeni sonori. Lo studio si è concentrato sull'analisi delle HRTF, e cioè sullo studio della funzione di trasferimento che lega l'evento sonoro alla testa dell'ascoltatore. Tale funzione non è altro che una misura della trasformazione che subisce l'evento sonoro dalla sorgente fino alle orecchie dell'ascoltatore, principalmente a causa della diffrazione con il corpo, con la testa e con i padiglioni auricolari.

L'analisi è stata fatta su una serie di dati ricavati attraverso una campagna di acquisizione, utilizzando due manichini binaurali e quattro soggetti reali.

La tesi è stata strutturata nel seguente modo. Nel primo capitolo viene fatta un'introduzione sulle modalità relative alla percezione sonora e la definizione delle HRTF. Nel secondo capitolo vengono descritti gli algoritmi di identificazione delle risposte impulsive utilizzati nel corso della campagna di acquisizione effettuata presso il laboratorio Audio DSP del Dipartimento di Ingegneria dell'Informazione. Nel terzo capitolo vengono descritte le modalità di acquisizione ed elaborazione dei dati, mentre nell'ultimo capitolo vengono commentati e confrontati i risultati ottenuti.

Indice

Elenco delle figure	iii
Acronimi	iv
1 Localizzazione del suono	1
1.1 Percezioni sonore interaurali	1
1.2 Percezioni sonore spettrali	3
1.3 Head-Related Transfer Function (HRTF)	4
1.3.1 Problematiche delle HRTF	4
2 Algoritmi di identificazione delle HRTF	6
2.1 Tecnica basata su MLS	6
2.2 Tecnica basata sul Codice di Golay	8
2.3 Tecnica basata sullo sweep	9
3 Acquisizione ed elaborazione dei dati	12
3.1 Acquisizione dei dati	12
3.2 Script di elaborazione dei dati	13
3.2.1 Implementazione dell'algoritmo MLS	15
3.2.2 Implementazione dell'algoritmo basato su Codice di Golay	16
3.2.3 Implementazione della Tecnica Sweep	16
3.3 Elaborazione dei dati	17
4 Analisi dei dati	18
4.1 Confronto tra le diverse tecniche nella configurazione Testa Busto (TB)	19
4.2 Confronto tra orecchio destro e orecchio sinistro nella configurazione Testa Busto (TB)	19
4.3 Confronto considerando la sorgente a diverse altezze nella configurazione TB	22
4.4 Analisi della simmetria tra l'orecchio destro e l'orecchio sinistro nella configurazione TB	24
4.5 Confronto della configurazione Testa Busto (TB) e Testa (T)	25
4.6 Valutazione delle risposte tra diversi soggetti umani	29
4.6.1 Variazione dell'altezza della sorgente	29

4.6.2	Analisi della simmetria tra l'orecchio destro e l'orecchio sinistro	32
	Conclusioni	36
	Bibliografia	38

Elenco delle figure

1.1	Ambiguità della ITD	2
1.2	Cono di confusione	3
1.3	ITD (sinistra) e ILD (destra)	4
2.1	Registro a scorrimento per la generazione di una sequenza MLS	7
2.2	Spettrogramma di uno sweep logaritmico	10
3.1	Soggetti utilizzati	13
3.2	Schema del setup di acquisizione dei dati	14
3.3	Immagine del setup di acquisizione dei dati	15
4.1	Confronto fra i tre metodi per l'orecchio sinistro 120°	20
4.2	Confronto fra i tre metodi per l'orecchio sinistro 100° in frequenza	21
4.3	Disposizione degli angoli	22
4.4	Configurazione TB (180° metodo MLS)	23
4.5	Configurazione TB nel tempo (100° metodo MLS)	24
4.6	Configurazione delle altezze	25
4.7	Configurazione TB (180° metodo sweep)	26
4.8	Confronto della simmetria (100° offset metodo Sweep)	27
4.9	Confronto configurazioni (260° metodo Sweep)	27
4.10	Confronto configurazioni (100° metodo mls)	28
4.11	Confronto configurazioni (180° metodo mls)	28
4.12	Posizionamento soggetto reale	29
4.13	Confronto delle altezze del soggetto P1	30
4.14	Confronto delle altezze del soggetto P2	31
4.15	Confronto delle altezze del soggetto P3	31
4.16	Confronto delle altezze del soggetto P4	32
4.17	Simmetria soggetti (80° offset metodo MLS)	33
4.18	Risposte in frequenza dei 4 soggetti a 140° per l'orecchio sinistro	34
4.19	Risposte in frequenza dei 4 soggetti a 180°	35

Acronimi

HRTF Head Related Transfer Function. 2, 4–6, 12, 13, 34, 36

ILD Interaural Level Difference. 1–4

IR Impulse Response. 4, 7–11, 14–16, 19, 37

ITD Interaural Time Difference. 1–4

T Testa. i, 12, 25

TB Testa Busto. i, iii, 12, 18, 19, 22–26

TF Transfer Function. 4, 19, 36

Capitolo 1

Localizzazione del suono

Si è consapevoli che l'uomo non solo riesce a percepire i suoni, ma anche a identificare con buona precisione la posizione di una sorgente sonora. Verranno ora analizzati gli aspetti che caratterizzano la percezione sonora e che sono alla base di questo importante meccanismo.

1.1 Percezioni sonore interaurali

All'inizio del XIX secolo il fisico britannico Rayleigh analizzò i principali meccanismi che entrano in gioco nella percezione della direzione di una sorgente sonora, ideando la "duplex theory" [1]. Questa teoria si basa sull'analisi di due grandezze:

- **Interaural Level Difference (ILD)** o differenza inter-aurale, definita come la differenza di ampiezza della sorgente sonora che arriva a ciascun orecchio. In particolare ha un maggior effetto per frequenze maggiori di 1.5 kHz perché l'effetto è principalmente introdotto dalla testa che fa da "ombra" all'orecchio più lontano, attenuando più facilmente le frequenze più alte.
- **Interaural Time Difference (ITD)** o differenza di ritardo inter-aurale, definita come la differenza temporale di arrivo e quindi di fase del segnale sonoro dalla sorgente alle due orecchie. In particolare ha un maggior effetto per fre-

quenze minori di 1.5 kHz, dato che a frequenze maggiori possono verificarsi delle ambiguità nella determinazione della ITD, come riportato in Figura 1.1.

Gli esseri umani mostrano notevoli capacità nella localizzazione del suono, riuscendo a discriminare la posizione di due sorgenti con circa $10 \mu s$ di differenza di ITD e tra 1 dB e 2 dB di differenza per la ILD [2]. Uno studio recente [3] ha determinato che l'accuratezza nell'identificare una sorgente si è dimostrato indipendente sia dal livello di volume della sorgente che dalla durata, mentre l'accuratezza della localizzazione è migliorata aumentando la larghezza di banda dello stimolo. Inoltre, sovrapponendo un rumore (nel caso di questo studio di tipo gaussiano) al segnale utile, si è visto un peggioramento nelle capacità di identificazione della posizione di tale sorgente se il disturbo proveniva da una posizione laterale rispetto ad una posizione frontale.

1. Below 1500 Hz ("low-frequency") 2. Above 1500 Hz ("high-frequency")

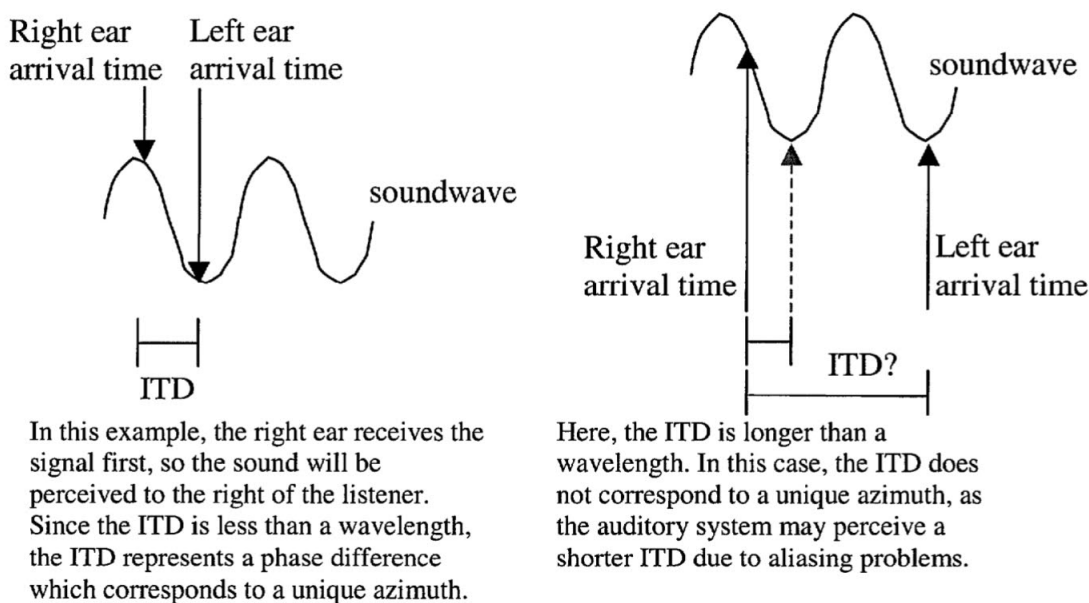


Figura 1.1: Ambiguità della ITD

Queste percezioni determinano l'identificazione della sorgente nel piano orizzontale, ma non bastano a identificarne l'elevazione, dando origine al così detto "cono di confusione", ovvero alla creazione di infinite posizioni in cui i parametri ITD e ILD sono uguali (Figura 1.2) [4].

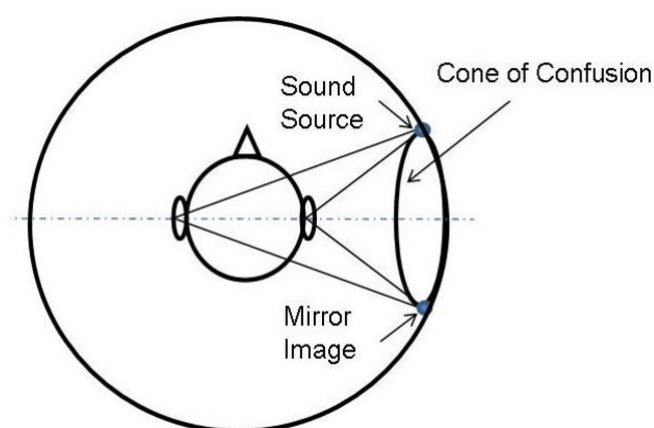


Figura 1.2: Cono di confusione

1.2 Percezioni sonore spettrali

Altre percezioni, come l'interazione dell'onda sonora con la testa dell'ascoltatore, il busto e il padiglione auricolare, danno una colorazione spettrale tale da permettere al cervello di identificare la posizione della sorgente sul piano verticale.

Uno studio condotto da Batteau Dwight W. [5], ipotizza che la percezione direzionale provenga dalle diverse riflessioni che si formano all'interno del padiglione auricolare, introducendo diversi ritardi che variano con la direzione e elevazione della sorgente.

Un altro studio riportato in [6] ha determinato che la direzione di una sorgente sonora può essere identificata con un solo orecchio, grazie anche ai risultati di un esperimento di Fisher e Freedman (1968) [7]. In particolare, in questo studio, i soggetti dovevano identificare la posizione di otto diffusori localizzati a 45° sul piano orizzontale rispetto all'ascoltatore. Dall'analisi dei risultati si è evinto che l'accuratezza delle localizzazioni era pressoché la stessa, sia con un orecchio occluso che con entrambe le orecchie scoperte. Successivamente, è stato inserito un tubo lungo 10 cm nel canale uditivo dell'ascoltatore, e si è dimostrato che l'accuratezza della localizzazione con un orecchio occluso veniva notevolmente ridotta, a causa dell'assenza delle riflessioni all'interno del padiglione auricolare.

Questi studi dimostrano che non solo esistono delle percezioni binaurali come ILD e ITD, ma esistono anche percezioni monoaurali, causate dalla colorazione spettrale

introdotta dal padiglione auricolare.

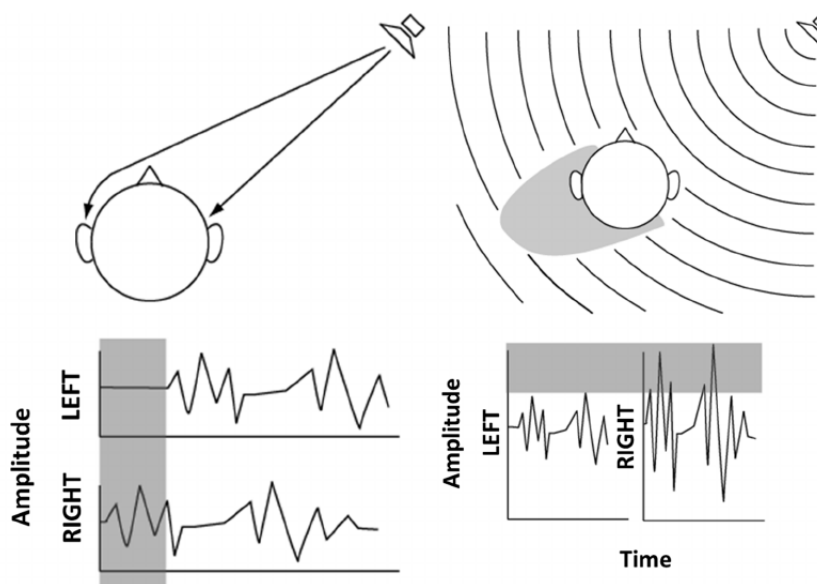


Figura 1.3: ITD (sinistra) e ILD (destra)

1.3 Head-Related Transfer Function (HRTF)

Come menzionato prima, l'interazione del suono con il nostro corpo modifica lo spettro della sorgente. La funzione di trasferimento relativa alla testa, o HRTF, è una coppia di funzioni di trasferimento o Transfer Function (TF), una per ciascun orecchio, e descrive la colorazione spettrale che il nostro padiglione auricolare dà ad una sorgente sonora. Queste variano non solo con la direzione orizzontale, ma anche con l'elevazione, la distanza e da persona a persona in modo complesso, a causa dei padiglioni auricolari di forme e grandezze diverse. I parametri ITD e ILD si possono visualizzare con la HRIR, ovvero la rispettiva risposta impulsiva o Impulse Response (IR) della HRTF [8].

1.3.1 Problematiche delle HRTF

Numerosi studi [4] [9] [10], hanno tentato di spazializzare diverse sorgenti sonore tramite l'utilizzo di calcolatori, per essere utilizzate ad esempio nei videogiochi, in

eventi musicali o nella realtà virtuale, ma risulta difficile utilizzare un set unico di HRTF che funzioni per qualsiasi ascoltatore basandosi su tecniche di sintesi binaurale. Data l'unicità del padiglione auricolare, ogni soggetto possiede una coppia unica di HRTF per una direzione ed un'elevazione specifica. Ascoltare sorgenti sonore filtrate dalle HRTF di un altro soggetto, può generare errori di localizzazione dell'elevazione o creare rovesciamenti fronte-retro della posizione della sorgente virtuale.

La difficoltà maggiore, nel tentativo di spazializzare sorgenti sonore in tempo reale, è per quelle in movimento, perché essendo a conoscenza di un set finito di HRTF (ogni set è unico per ogni direzione ed elevazione), bisogna fare in modo che la transizione da una HRTF all'altra avvenga in modo continuo, richiedendo quindi una interpolazione fra loro, e questo in tempo reale necessita una grande potenza di calcolo. Numerosi studi [4], [9] sostengono che bisogna capire più a fondo gli elementi principali che caratterizzano le HRTF, soprattutto per l'effetto dell'elevazione, in modo tale da trovare algoritmi più efficienti e quindi un costo computazionale minore, e cercare di migliorare le performance delle HRTF non individualizzate.

Capitolo 2

Algoritmi di identificazione delle HRTF

Per ricavare le coppie di risposte impulsive e le HRTF, sono state utilizzate delle sequenze con delle proprietà matematiche particolari, che hanno permesso di estrarre facilmente le risposte con un rapporto segnale rumore sufficientemente alto. Le misurazioni sono state fatte comparando tre metodi diversi:

- La tecnica basata sulla sequenza MLS (Maximum Length Sequences);
- La tecnica basata sul Codice di Golay;
- La tecnica basata sullo sweep logaritmico;

Nel corso delle misurazioni, il sistema tra la sorgente acustica e il ricevitore è stato considerato come lineare a tempo invariante (LTI).

2.1 Tecnica basata su MLS

Una Maximum Length Sequence (MLS) [11] è una sequenza binaria⁽¹⁾ pseudorandomica periodica di lunghezza infinita (con energia del segnale uniformemente distribuita nel periodo), generata da un registro a scorrimento a retroazione lineare detto anche

⁽¹⁾Con questo metodo si ottiene una sequenza unipolare composta da 0 e 1, ma si può facilmente trasformarla in una sequenza bipolare composta da 1 e -1.

”Linear Feedback Shift Register” o LFSR. Il numero di registri L , utilizzato per generare la sequenza MLS, determina la lunghezza del periodo $N = 2^L - 1$ (la sequenza con tutti 0 è esclusa come condizione iniziale dei registri).

Il segnale viene generato sotto il controllo di un clock, e ad ogni colpo di clock lo stato di ogni registro (0 o 1) viene traslato o ”shiftato” in quello successivo, quindi si ha che:

$$\begin{aligned} a_i(n+1) &= a_{i-1}(n), \quad \text{per } i \in \{2, 3, \dots, L\}, \\ a_1(n+1) &= c_1 a_1(n) \oplus c_2 a_2(n) \oplus \dots \oplus c_{L-1} a_{L-1}(n) \oplus c_L a_L(n) \end{aligned} \quad (2.1)$$

dove $n+1$ rappresenta il colpo di clock successivo a n , il simbolo \oplus denota l’operatore logico ”xor”, mentre c_i rappresenta il coefficiente di retroazione, che può essere 0 o 1, eccetto c_L ($c_L = 1$). In Figura 2.1 si vede una rappresentazione del registro a scorrimento.

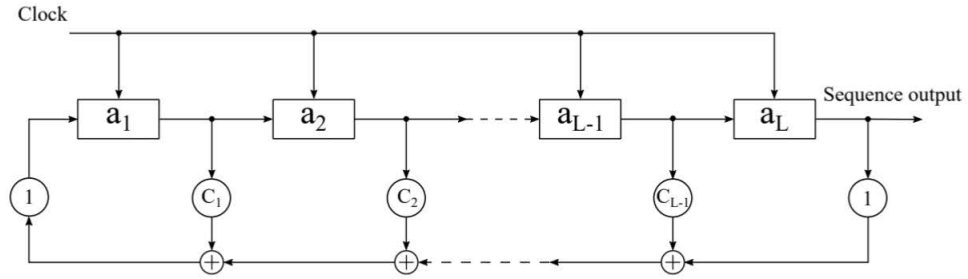


Figura 2.1: Registro a scorrimento per la generazione di una sequenza MLS

Supponendo $s[n]$ una sequenza MLS, da questa sequenza si ricava la sua autocorrelazione circolare nel seguente modo:

$$R_{ss}[n] = \sum_{k=0}^{N-1} s[k] \cdot s[n+k]_N = \begin{cases} N, & n = 0 \\ -1, & n \neq 0 \end{cases} . \quad (2.2)$$

Dividendo il risultato dell’autocorrelazione per N , e supponendo che questo sia relativamente grande, si può approssimare una delta di Kroneker. Questo significa che è possibile ricavare una IR sfruttando questa proprietà. Supponendo che $s[n]$ sia il segnale originale, mentre $c[n] = s[n] \otimes h[n]$ sia il segnale registrato, con $h[n]$ la IR del

sistema:

$$\begin{aligned} \frac{1}{N} \cdot \text{xcorr}(c[n], s[n]) &= \frac{1}{N}(c[n] \otimes s[-n]) = \frac{1}{N}(s[n] \otimes s[-n] \otimes h[n]) \approx \\ &\approx \frac{1}{N}(N \cdot \delta[n] \otimes h[n]) = h[n] \end{aligned} \quad (2.3)$$

dove il simbolo \otimes sta ad indicare l'operatore di convoluzione.

I vantaggi di questa sequenza sono la semplicità di implementazione e l'utilizzo di uno stimolo con energia uniformemente distribuita, un vantaggio rispetto ad un semplice impulso la cui energia è distribuita maggiormente all'inizio, rendendolo vulnerabile al rumore quando questo si presenta nella coda dell'impulso.

Uno degli svantaggi della MLS è che idealmente il segnale è periodico e infinito, ed è quindi impossibile riprodurre il segnale per intero, di conseguenza non si otterrà esattamente la IR, ma una sua stima. Inoltre, per evitare problemi di aliasing, dato l'utilizzo della correlazione circolare, l'ordine della sequenza deve essere abbastanza grande da avere una lunghezza maggiore della IR. Un altro svantaggio che può presentarsi utilizzando la MLS [12] è la scelta dell'ordine di questa, dato che un ordine più alto diminuisce il livello del rumore che si sovrappone alla IR, ma aumenta anche la durata della sequenza, rendendola vulnerabile a fenomeni tempo varianti.

Questo tipo di sequenza non ha un'alta immunità a distorsioni di non linearità, infatti uno studio di Dunn Chris e Hawksford Malcolm [13] ha dimostrato come questi disturbi vengono distribuiti uniformemente nella IR, e possono essere notevolmente ridotti in una versione modificata del MLS, ovvero utilizzando la sequenza IRS (Inverse Repeat Sequence), che pur mantenendo le proprietà del segnale MLS, mostra un'alta immunità contro i disturbi non lineari.

2.2 Tecnica basata sul Codice di Golay

I codici di Golay sono coppie di segnali binari deterministici $a[n]$ e $b[n]$ della stessa lunghezza L , definiti dalla seguente proprietà [14]:

$$\begin{aligned} R_{aa}[n] + R_{bb}[n] &= a[n] \otimes a[-n] + b[n] \otimes b[-n] \\ &= 2A^2L \cdot \delta[n] \end{aligned} \quad (2.4)$$

dove $a[n]$ e $b[n]$ valgono $+A$ e $-A$ per $n \in [0, L - 1]$, nulla altrove. Nel caso dell'identificazione della IR di un sistema, si considerano $a_0[n]$ e $b_0[n]$ i segnali di partenza, $a_1[n]$ e $b_1[n]$ i segnali registrati, la IR vale:

$$h[n] = \frac{R_{a_0 a_1}[n] + R_{b_0 b_1}[n]}{2A^2L}. \quad (2.5)$$

Sebbene questi codici esistono di lunghezze differenti, verrà trattata solo la classe dei codici di Golay la cui lunghezza sia una potenza esatta del 2, cosicché $L = 2^p$, con p l'ordine della sequenza.

Esistono diversi algoritmi per costruire questi codici: Golay ha fornito un semplice algoritmo di negazione e concatenazione che fornisce questa classe di codici. La definizione, ricorsiva, è la seguente: se \mathbf{a} e \mathbf{b} sono sequenze binarie complementari di lunghezza L , allora la nuova coppia di sequenze $\mathbf{a1}$, $\mathbf{b1}$ di lunghezza $2L$ può essere generato concatenando \mathbf{b} ad \mathbf{a} per ottenere $\mathbf{a1}$ e concatenando $\mathbf{not\ b}$ ad \mathbf{a} per ottenere $\mathbf{b1}$. Dunque $\mathbf{a1}=[\mathbf{a} \mid \mathbf{b}]$ e $\mathbf{b1}=[\mathbf{a} \mid \mathbf{not\ b}]$.

Il principale vantaggio sta nel fatto che i segnali sono di lunghezza finita e ritornano precisamente la IR. Lo svantaggio è la procedura di implementazione più complicata, visto che si devono riprodurre due segnali per identificare il sistema anziché uno solo. Se il sistema è suscettibile a cambiamenti durante la presentazione dei due segnali, ci potrebbero essere errori significativi nella stima della risposta.

2.3 Tecnica basata sullo sweep

Lo sweep è un segnale sinusoidale la cui frequenza cresce all'aumentare del tempo, del tipo:

$$x(t) = A \cdot \sin(\varphi(t)), \quad e \quad f(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt} \quad (2.6)$$

dove A è l'ampiezza del segnale, $\varphi(t)$ e $f(t)$ sono rispettivamente la fase istantanea e la frequenza.

Nel corso delle misurazioni è stato utilizzato uno sweep di tipo logaritmico (spesso chiamato anche sweep esponenziale) da 100 Hz (f_1) a 20 kHz (f_2) della durata di 5 secondi (T), in Figura 2.2 si vede lo spettrogramma di tale segnale, mentre la

frequenza è del tipo:

$$f(t) = f_1 e^{ct}, \quad \text{dove} \quad c = \frac{\ln(\frac{f_2}{f_1})}{T} \quad (2.7)$$

e dopo aver ricavato la fase istantanea si ottiene lo sweep logaritmico $x(t)$:

$$x(t) = A \cdot \sin\left(2\pi \frac{f_1 T}{\ln(\frac{f_2}{f_1})} \left(e^{\frac{t}{T} \ln(\frac{f_2}{f_1})} - 1\right)\right). \quad (2.8)$$

Con questo metodo si recupera in modo accurato la IR solo per l'intervallo di frequenze definito nello sweep.

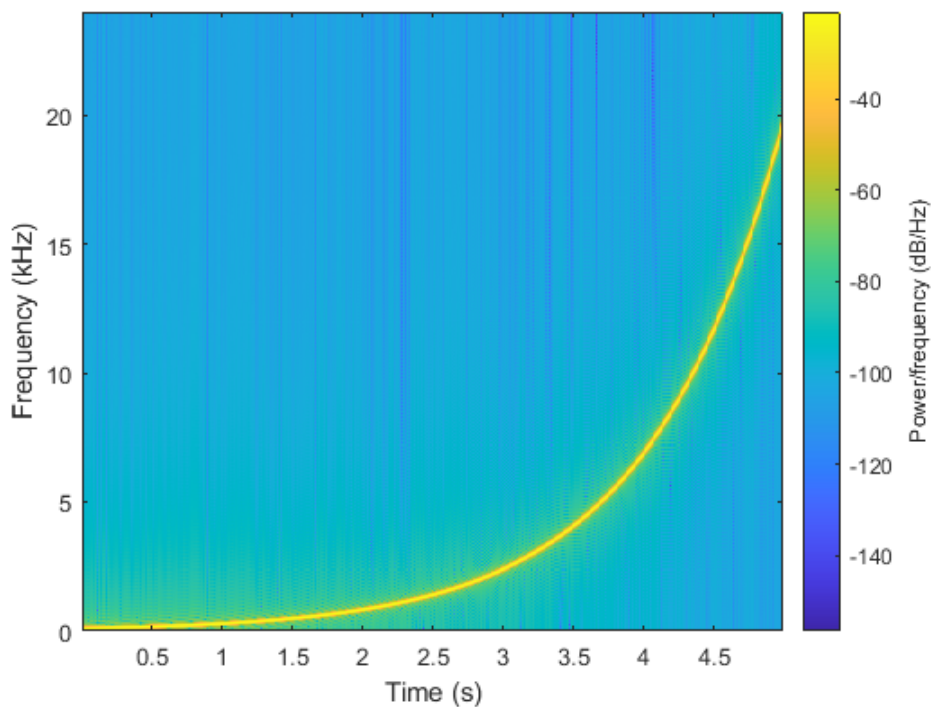


Figura 2.2: Spettrogramma di uno sweep logaritmico

Dato che il sistema di analisi viene considerato come lineare a tempo invariante, si sfrutta la seguente proprietà per ricavare la risposta in frequenza:

$$H(f) = \frac{Y(f)}{X(f)} \quad (2.9)$$

dove $H(f)$ è la FFT⁽²⁾ della IR $h[n]$, $Y(f)$ è la FFT del segnale registrato e $X(f)$ è la FFT dello sweep di partenza.

⁽²⁾Fast Fourier Transform

In letteratura [12], questo metodo di misurazione della IR è quello più utilizzato, data la sua semplicità di implementazione, e dal fatto che eventuali armoniche che possono generarsi a causa di comportamenti non lineari del sistema misurato, non si sovrappongono alla IR, a differenza del metodo MLS dove il rumore viene uniformemente distribuito su tutta la IR. Uno dei pochi svantaggi [15] è la vulnerabilità a rumori di tipo impulsivo, e può essere utilizzato solamente in ambienti silenziosi.

Capitolo 3

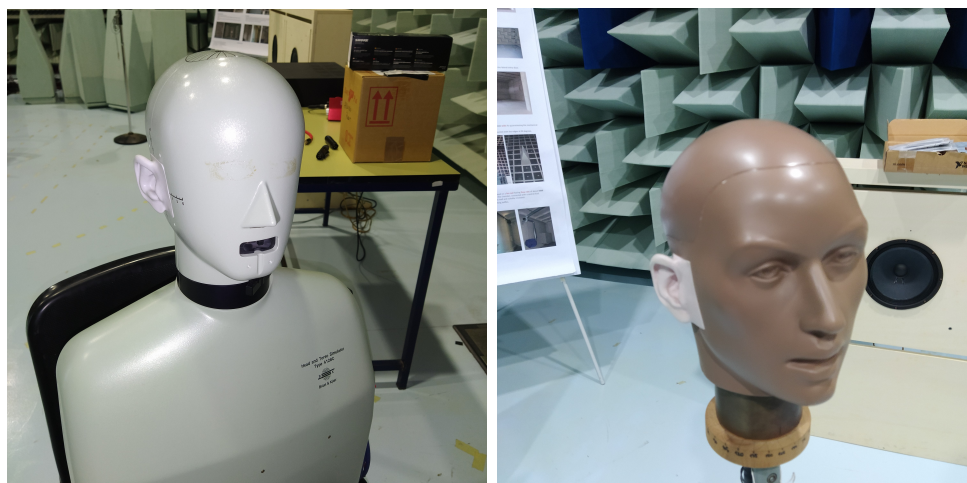
Acquisizione ed elaborazione dei dati

3.1 Acquisizione dei dati

Per l'acquisizione delle HRTF è stata utilizzata la camera semianecoica del Dipartimento di Ingegneria dell'Informazione (DII) dell'Università Politecnica della Marche. Per le prove è stato utilizzato un monitor da studio Genelec 8020A collegato ad un canale di output della scheda audio Focusrite Scarlett 2i2, collegato ad un PC portatile (AMD A8 a 2 Ghz) sul quale è stato utilizzato il software Nu-Tech⁽¹⁾, che permette di implementare algoritmi DSP in tempo reale a bassa latenza, per riprodurre e registrare le sequenze contemporaneamente in automatico.

Ai due canali di input della scheda audio sono stati collegati i due microfoni (una per ciascun orecchio) delle configurazioni utilizzate, inizialmente nella configurazione Testa Busto (TB) con il manichino della B&K modello 4128C, poi solo Testa (T) con il manichino KEMAR con accoppiatori di Zwislocki, un dispositivo che simula le caratteristiche del canale uditivo, e infine, utilizzando dei microfoni interaurali "Sennheiser mke 2-ew gold" posti all'ingresso del canale uditivo di quattro soggetti. Tra il microfono e il canale di input della scheda audio è stato inserito un preamplificatore.

⁽¹⁾<http://www.nu-tech-dsp.com/>



(a) Manichino B&K

(b) Testa del Kemar



(c) Microfoni interaurali

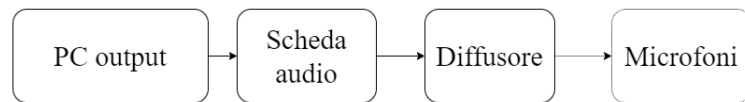
Figura 3.1: Soggetti utilizzati

3.2 Script di elaborazione dei dati

Tramite il programma MATLAB sono stati implementati dei codici di elaborazione dei dati registrati, per ricavare le HRTF e le HRIR. Tutta la documentazione relativa alle funzioni utilizzate per l'implementazione del codice si trova direttamente nel sito web⁽²⁾ di MATLAB. Ogni funzione elabora la risposta di un solo orecchio per volta. Ognuna di queste inizia con le seguenti linee di codice:

⁽²⁾<https://it.mathworks.com/help/matlab/>

Fase di output



Fase di input

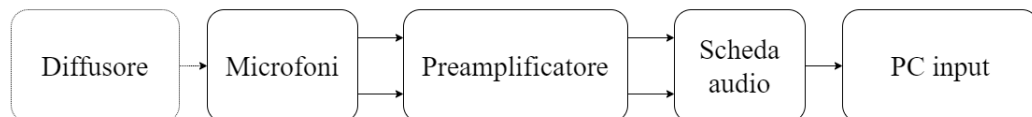


Figura 3.2: Schema del setup di acquisizione dei dati

```

1 mic=audioread(fileMic);
2 orig=audioread(fileOrig);
  
```

Nel caso del codice di Golay in cui si utilizzano due segnali:

```

1 micA=audioread(fileMicA);
2 micB=audioread(fileMicB);
3 origA=audioread(fileOrigA);
4 origB=audioread(fileOrigB);
  
```

Si parte dalla funzione `audioread(file)` che prende in ingresso la directory del file .wav del segnale di partenza (variabile `orig`) e del segnale registrato da uno dei due microfoni (variabile `mic`), mettendone il contenuto all'interno delle rispettive variabili. Successivamente, vengono calcolate le risposte sfruttando le proprietà di ciascun metodo. In tutte le funzioni implementate, le IR ricavate vengono finestrate temporalmente, eliminando le latenze di input e output introdotte dal Nu-Tech, con una lunghezza di 4096 campioni. Infine, queste vengono salvate sotto forma di file .dat tramite le seguenti linee di codice:



Figura 3.3: Immagine del setup di acquisizione dei dati

```

1 f=fopen(name, 'w');
2 fwrite(f, IR, 'float64');
3 fclose(f);

```

Per leggere il file .dat della IR:

```

1 fid=fopen(name, 'r');
2 IR=fread(fid, 'float64');
3 fclose(fid);

```

3.2.1 Implementazione dell'algoritmo MLS

Per calcolare la IR con metodo MLS è stata utilizzata la formula 2.3:

```

1 L=length(orig);
2 h=xcorr(mic,orig)/L;
3
4 %punto di inizio in cui considerare la risposta impulsiva
5 inizio=floor((length(h)/2))+delay;
6 hLen=4096;
7 h=h(inizio:inizio+hLen);

```

3.2.2 Implementazione dell'algoritmo basato su Codice di Golay

Per calcolare la IR con il codice di Golay è stata utilizzata la formula 2.5:

```

1 L=length(origA);
2 %La variabile taglio viene definita in modo da avere i segnali ...
   registrati dal microfono della stessa lunghezza
3 taglio=min(length(micA),length(micB));
4 hA=xcorr(micA(1:taglio),origA);
5 hB=xcorr(micB(1:taglio),origB);
6 h=(hB+hA)/(2*L);
7 %punto di inizio in cui considerare la risposta impulsiva
8 inizio=floor((length(h)/2))+delay;
9 hLen=4096;
10 h=h(inizio:inizio+hLen);

```

3.2.3 Implementazione della Tecnica Sweep

Per calcolare la IR con metodo sweep è stata utilizzata la formula 2.9:

```

1 L=length(orig);
2 orig=orig(1:L);
3 micFreq=fft(mic(1+delay:L+delay));

```

```

4 origFreq=fft(orig);
5
6 h=ifft(micFreq./origFreq);
7 hLen=4096;
8 h=h(1:1+hLen);

```

3.3 Elaborazione dei dati

Per tutte le misurazioni sono stati utilizzati una sequenza MLS e una sequenza sweep di tipo logaritmico, mentre il codice di Golay è stato utilizzato solamente nei due manichini binaurali. Tutti i segnali sono in formato Wave codificati PCM a 16 bit con frequenza di campionamento di 48 kHz. Per il codice di Golay è stata scelta una sequenza di lunghezza 2^{18} per un totale di 262144 campioni, generata tramite seme generativo Golay 00. Per la sequenza MLS si è utilizzata una sequenza di ordine 18, con periodo $L = 2^{18} - 1$. Lo sweep logaritmico è stato creato tramite la funzione `chirp()` di MATLAB, lungo 5 secondi, partendo da una frequenza di 100 Hz fino a 20 kHz, in modo tale da coprire l'intero campo uditivo dell'orecchio umano (la banda 20 kHz - 24 kHz delle risposte in frequenza non ha valore significativamente utile).

```

1 t = 0:1/48000:5-1/48000;           % 5 s a 48 kHz sample rate
2 fo = 100;                          % parte da 100 Hz,
3 f1 = 20000;                         % fino a 20000 Hz
4 y = chirp(t,fo,10,f1,'logarithmic');

```

Capitolo 4

Analisi dei dati

Dopo aver posto in asse il manichino o il soggetto con il diffusore, posizionandolo ad una distanza di 1.5 m, sono state effettuate misurazioni con differenti configurazioni. Utilizzando i dati ottenuti dalle registrazioni in camera semianecoica, sono stati effettuati i seguenti confronti:

- Confronto tra le diverse tecniche nella configurazione Testa Busto (TB);
- Confronto tra orecchio destro e orecchio sinistro nella configurazione Testa Busto (TB);
- Confronto considerando la sorgente a diverse altezze nella configurazione TB;
- Analisi della simmetria tra l'orecchio destro e l'orecchio sinistro nella configurazione TB;
- Confronto della configurazione Testa Busto (TB) e Testa (T);
- Valutazione delle risposte tra diversi soggetti umani;

L'analisi in frequenza è stata fatta considerando la banda 100 Hz - 20 kHz. Nel dominio del tempo le ascisse indicano il numero di campioni, mentre le ordinate l'ampiezza. Nel dominio delle frequenze le ascisse indicano la frequenza in Hz, mentre le ordinate indicano le ampiezze in dB.

4.1 Confronto tra le diverse tecniche nella configurazione Testa Busto (TB)

Per fare un confronto tra tutte le tecniche usate per l'identificazione, verranno utilizzate le risposte del manichino B&K 4128C, dato che le altre configurazioni (i.e., manichino Kemar) hanno dato risultati analoghi.

Si vede dalla Figura 4.1 come i tre metodi siano comparabili fino ai 20 kHz, dato che il metodo sweep estrae correttamente la risposta in frequenza da 100 Hz a 20 kHz. Soltanto in alcuni casi è capitato che, nell'analisi delle risposte del codice di Golay, comparissero dei "notch" sempre nelle stesse frequenze, intorno ai 4 kHz, 12 kHz e 20 kHz. I notch sono visibili in Figura 4.2.

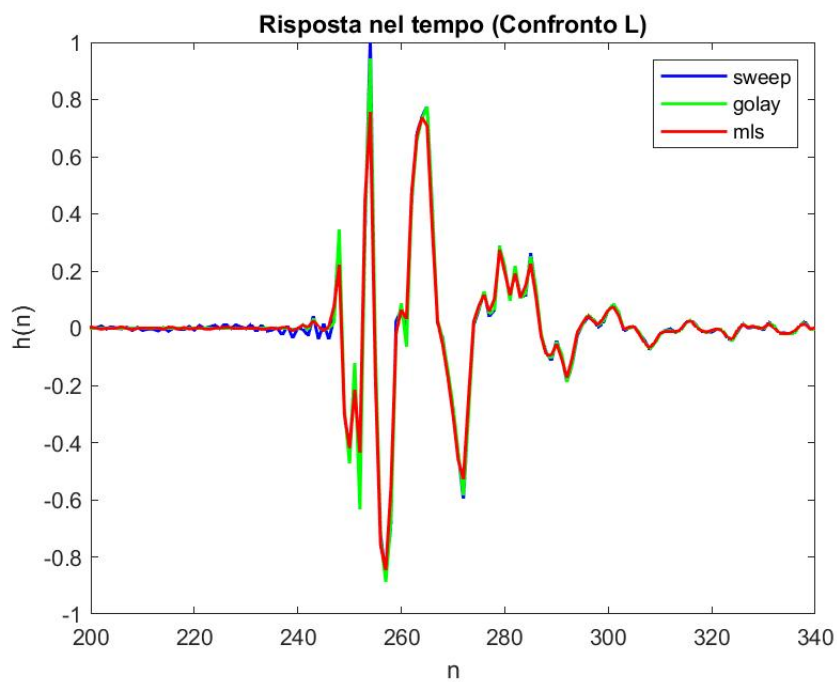
4.2 Confronto tra orecchio destro e orecchio sinistro nella configurazione Testa Busto (TB)

La configurazione di misura che si compone di Testa Busto (TB), ovvero quella del manichino B&K 4128C, è quella che meglio approssima la situazione reale umana di ascolto. Per questa configurazione sono stati sviluppati i grafici in frequenza e nel tempo, per confrontare la risposta dell'orecchio destro e dell'orecchio sinistro.

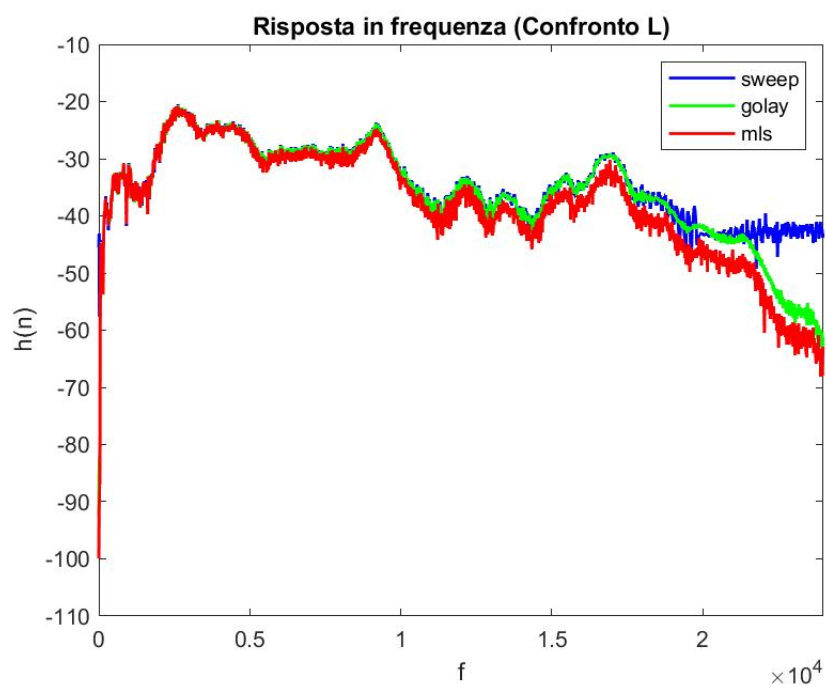
In fase di post processing, le curve nel dominio del tempo sono state normalizzate rispetto al valore massimo di picco di ogni IR. Per fare una valutazione sulla somiglianza tra l'orecchio destro e sinistro, l'analisi è partita da una posizione frontale al manichino (in Figura 4.3 viene mostrata la disposizione degli angoli).

Dalla Figura 4.4 si notano dissimmetrie nelle TF, anche se piuttosto simili, causate dalle forme disuguali dei padiglioni auricolari. Questo spiega l'importanza della sua struttura esterna che si comporta da filtro, modificando maggiormente l'andamento alle alte frequenze.

Si passa ora all'andamento nel tempo nel caso di un angolo di 100°. In Figura 4.5 è evidente come il segnale proveniente dall'orecchio destro sia in ritardo rispetto a



(a) Tempo



(b) Frequenza

Figura 4.1: Confronto fra i tre metodi per l'orecchio sinistro 120°

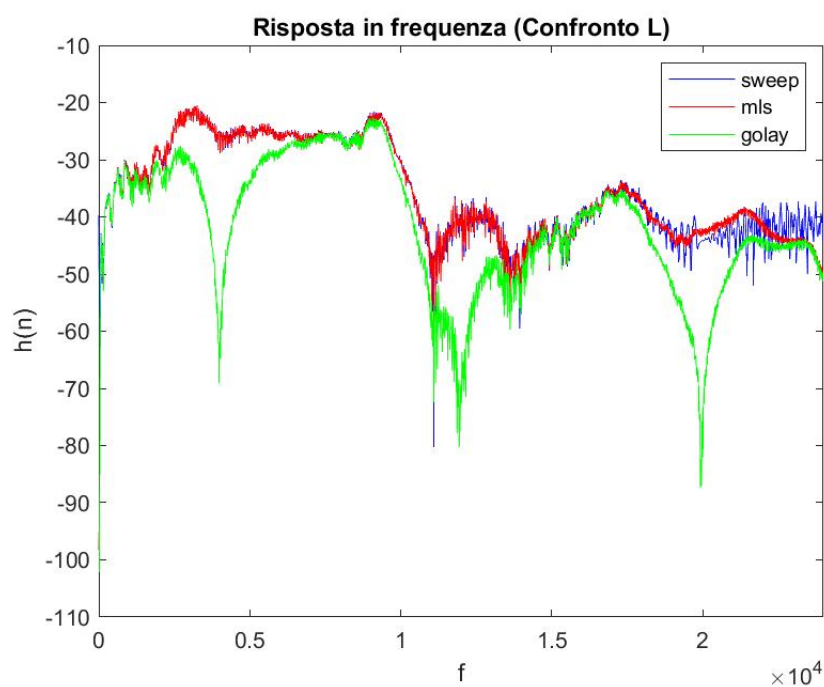


Figura 4.2: Confronto fra i tre metodi per l'orecchio sinistro 100° in frequenza

quello sinistro, dato che la sorgente proviene dalla sinistra del manichino. Infatti, se si contano i campioni ritardo, che si aggirano intorno ai 30 (questo ritardo è rimasto costante anche nelle altre configurazioni considerando lo stesso angolo), e tenendo conto della frequenza di campionamento a 48 kHz e la velocità del suono di 343 m/s, si calcola che la distanza in più percorsa dal suono per arrivare all'orecchio più lontano è di circa 20 cm, ovvero poco più della distanza tra le due orecchie, dato che il percorso che deve percorrere la sorgente per arrivare all'orecchio "nascosto" non è rettilineo e ma deve circondare la testa. Inoltre, l'ampiezza del segnale destro risulta nettamente minore rispetto a quello sinistro, a causa dell'effetto ombra che fa la testa sull'orecchio destro.

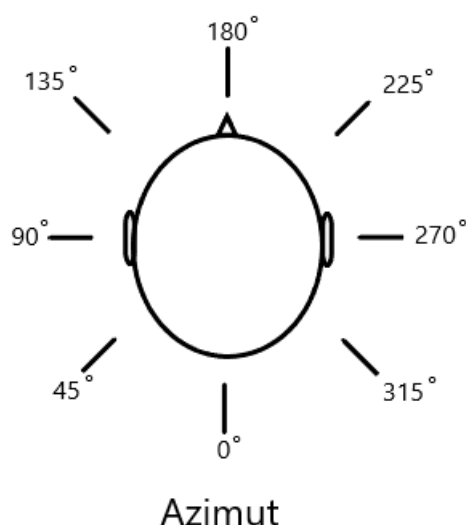


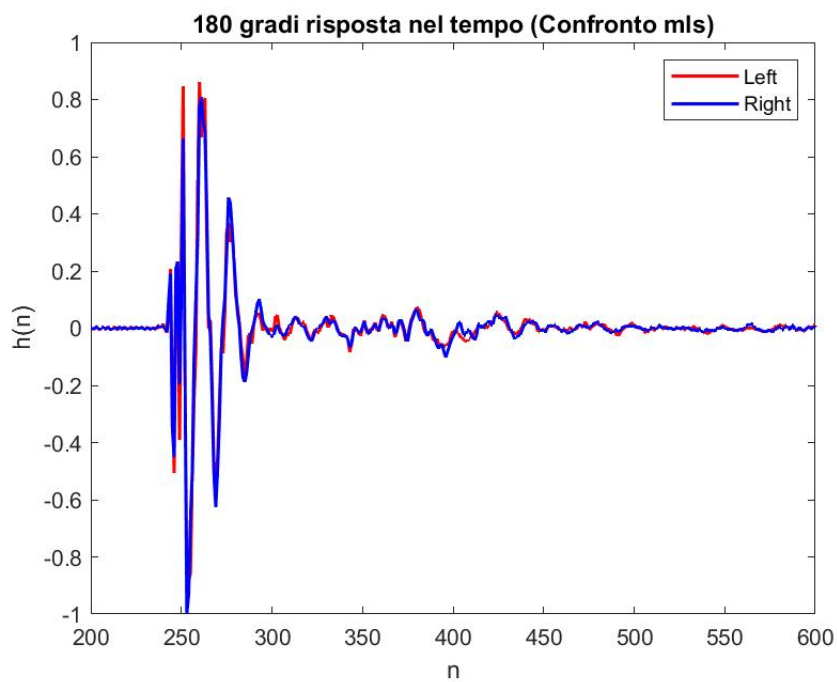
Figura 4.3: Disposizione degli angoli

4.3 Confronto considerando la sorgente a diverse altezze nella configurazione TB

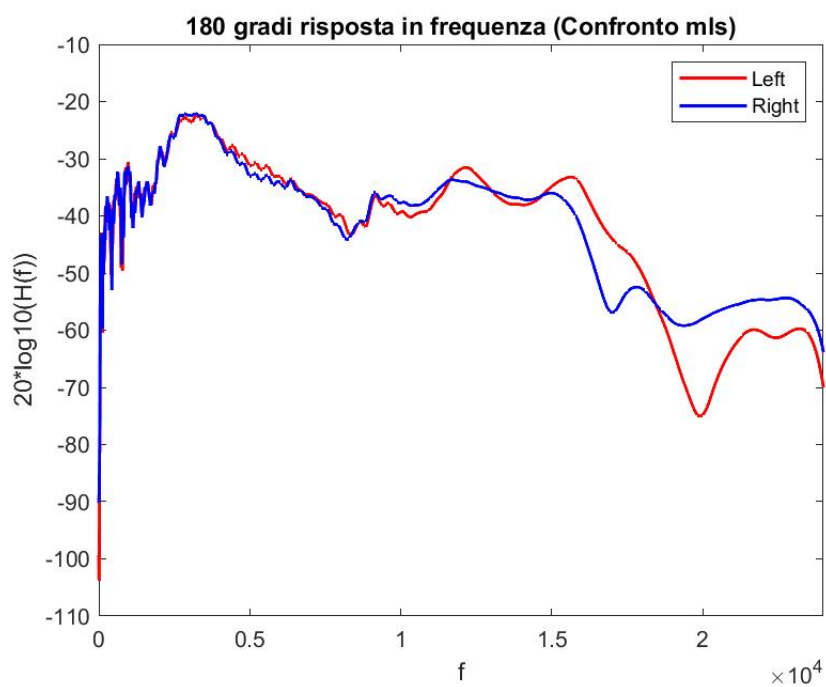
Ora viene analizzato il cambiamento delle risposte variando l'elevazione della sorgente. Nel caso della configurazione TB è stata elevata la sorgente di 7 cm e di 31 cm rispetto a quella di partenza che si trovava alla stessa altezza delle orecchie del manichino (Figura 4.6). Se si volesse esprimere l'elevazione in gradi, considerando la distanza tra sorgente e manichino di 150 cm, sono rispettivamente di circa 2.6° e 12° . Gli angoli azimutali in cui sono state fatte tali analisi sono 120° , 180° e 240° (Figura 4.3).

In Figura 4.7, a 180° si vede che, aumentando l'altezza, il picco presente a 15 kHz rimane invariato, e che a 120° e 240° nei range di frequenza tra circa 3 kHz e 7 kHz si presentano delle attenuazioni più evidenti quando l'elevazione è di 31 cm, mentre intorno agli 11 kHz il piccolo notch presente a 0 cm e 7 cm tende a scomparire elevando la sorgente a 31 cm.

Si vede chiaramente che il modo in cui le risposte cambiano con l'altezza, a 120° per l'orecchio sinistro e a 240° per l'orecchio destro, è piuttosto simile, in particolare fino ai 10 kHz.



(a) Tempo



(b) Frequenza

Figura 4.4: Configurazione TB (180° metodo MLS)

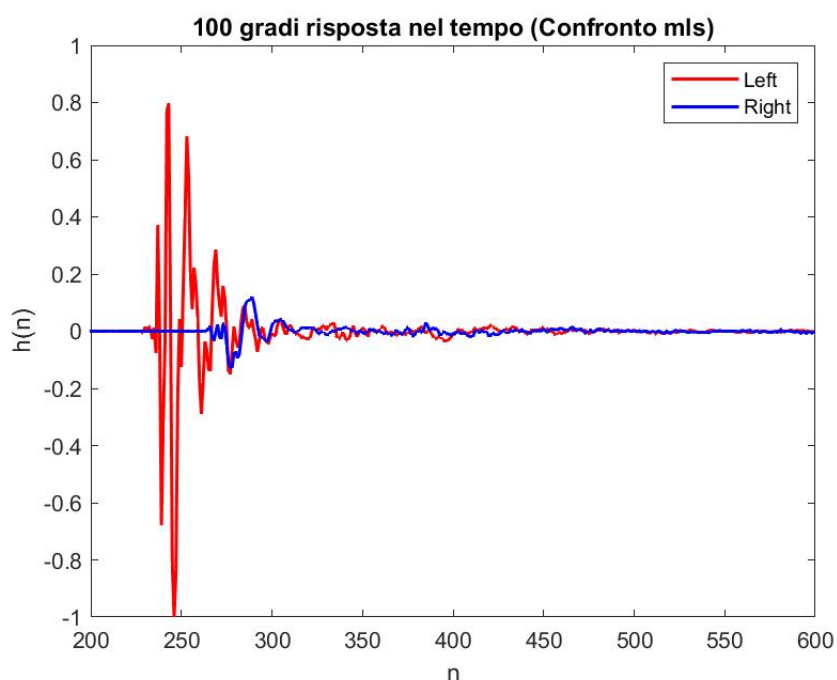
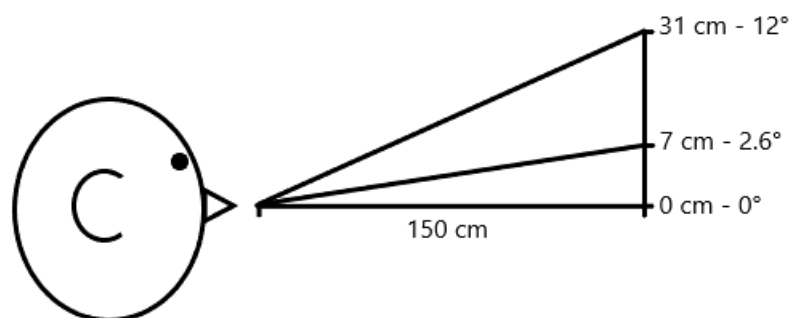


Figura 4.5: Configurazione TB nel tempo (100° metodo MLS)

4.4 Analisi della simmetria tra l'orecchio destro e l'orecchio sinistro nella configurazione TB

Si è già visto dalla Figura 4.4 che le risposte delle due orecchie non sono esattamente uguali. È possibile analizzare ulteriormente tale effetto andando ad inserire un angolo di offset rispetto al caso frontale, ad esempio, confrontando la risposta dell'orecchio sinistro a 80° con quella dell'orecchio destro a 280° (offset di 100°). In Figura 4.8 si vede che a partire da 10 kHz le risposte continuano con andamenti diversi.

L'effetto è tanto più accentuato quanto maggiore è l'offset, poiché i padiglioni auricolari possono presentare un angolo di sporgenza diverso, e questo influisce maggiormente quando il suono proviene da dietro.



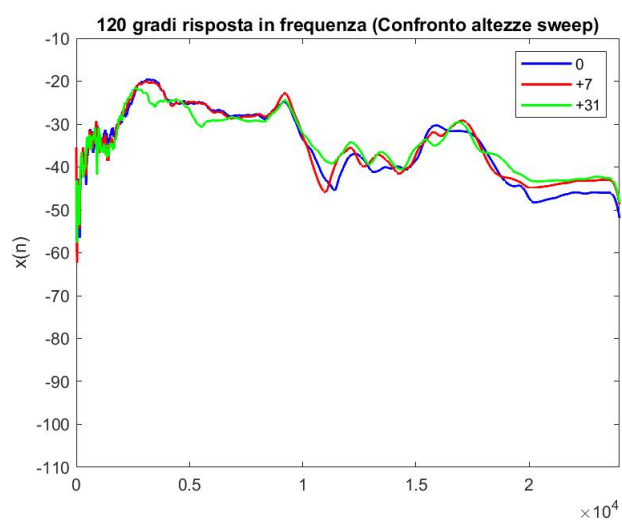
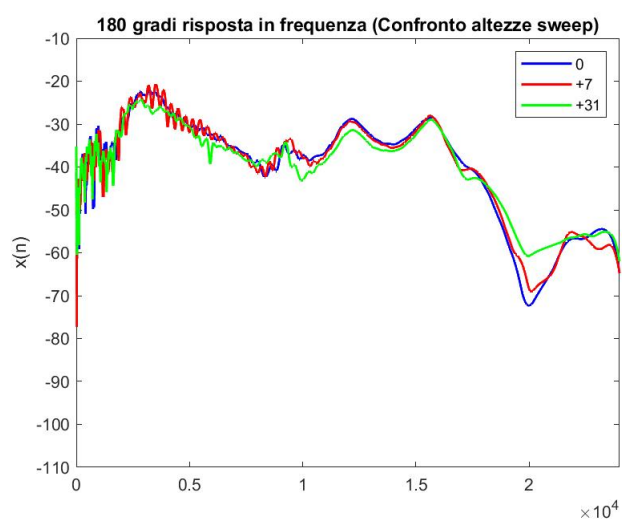
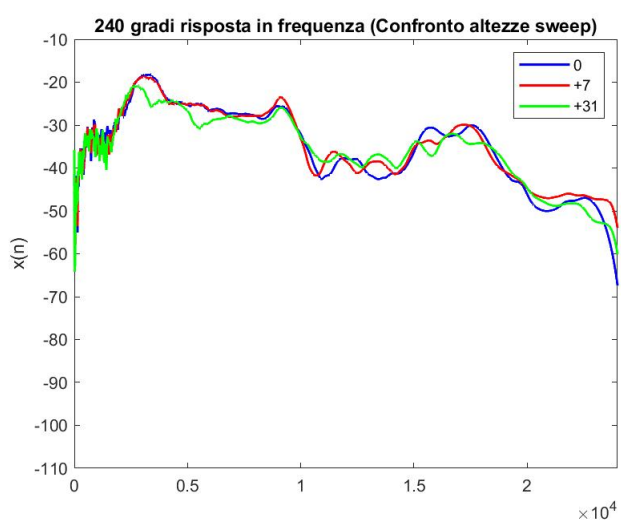
Elevazione

Figura 4.6: Configurazione delle altezze

4.5 Confronto della configurazione Testa Busto (TB) e Testa (T)

In fase di registrazione sono stati utilizzati due manichini differenti: uno composto da testa e busto, il manichino B&K modello 4128C, con i microfoni posti subito dopo il padiglione auricolare, l'altro, il manichino KEMAR, composto solo da testa con accoppiatori di Zwislocki. Come risultato ci si dovrebbe aspettare che la presenza del busto possa assorbire maggiormente l'onda sonora alle alte frequenze, ottenendo un'attenuazione più grande rispetto al modello con solo testa. In Figura 4.9 si vede molto bene l'effetto attenuante a partire da circa 10 kHz. Inoltre, le risposte in frequenza hanno andamenti simili ma non uguali, a causa dell'assenza del busto in una delle configurazioni. Oltre al comportamento attenuante alle alte frequenze, gli andamenti alle basse frequenze si somigliano ma i picchi e notch nella configurazione TB sono spostati di poco a frequenze più alte rispetto alla configurazione T. Un andamento molto simile è visibile a 100° in Figura 4.10.

Se si considera invece l'angolo frontale a 180°, si notano solamente delle attenuazioni alle alte frequenze per la configurazione T, come in Figura 4.11.

(a) 120° (b) 180° (c) 240° Figura 4.7: Configurazione TB (180° metodo sweep)

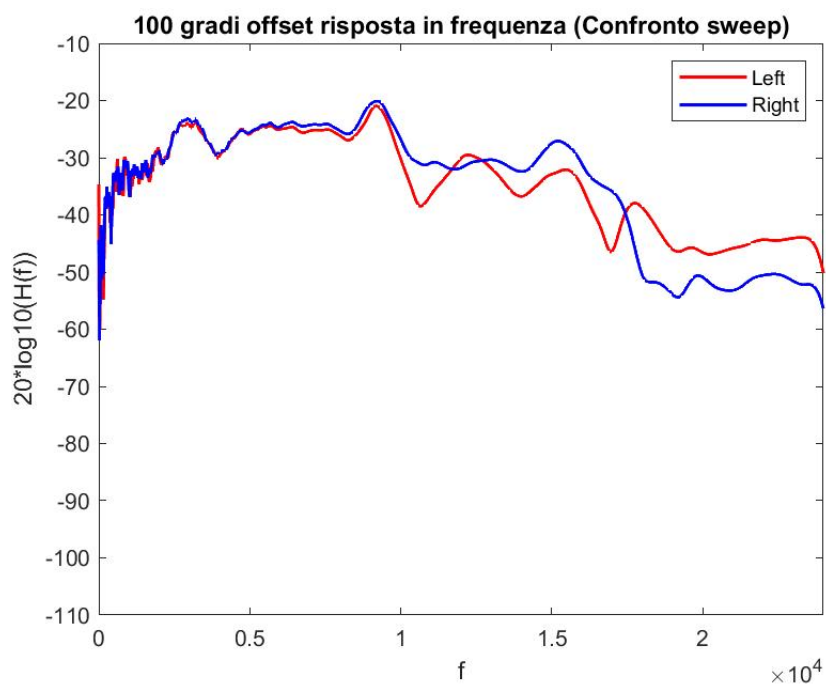


Figura 4.8: Confronto della simmetria (100° offset metodo Sweep)

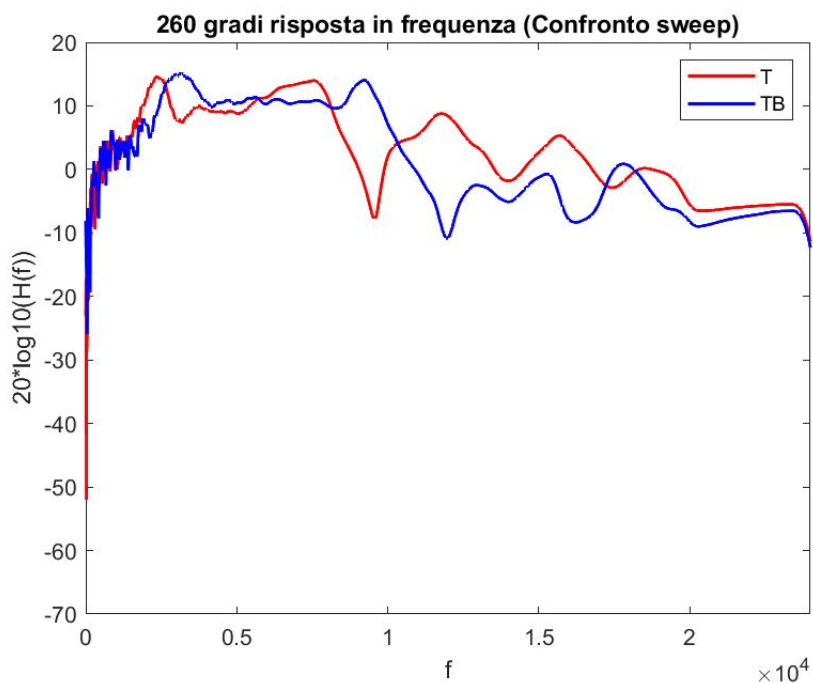


Figura 4.9: Confronto configurazioni (260° metodo Sweep)

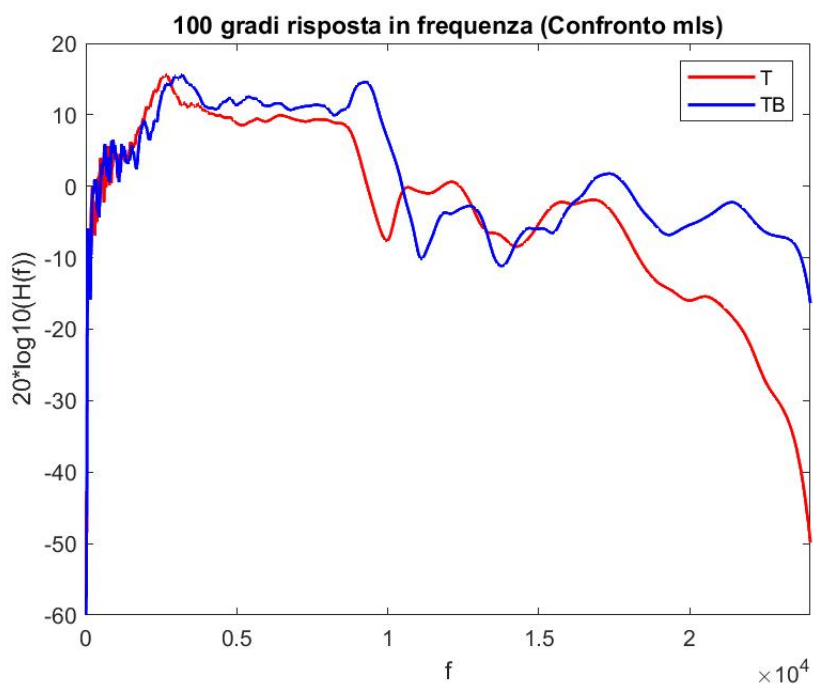


Figura 4.10: Confronto configurazioni (100° metodo mls)

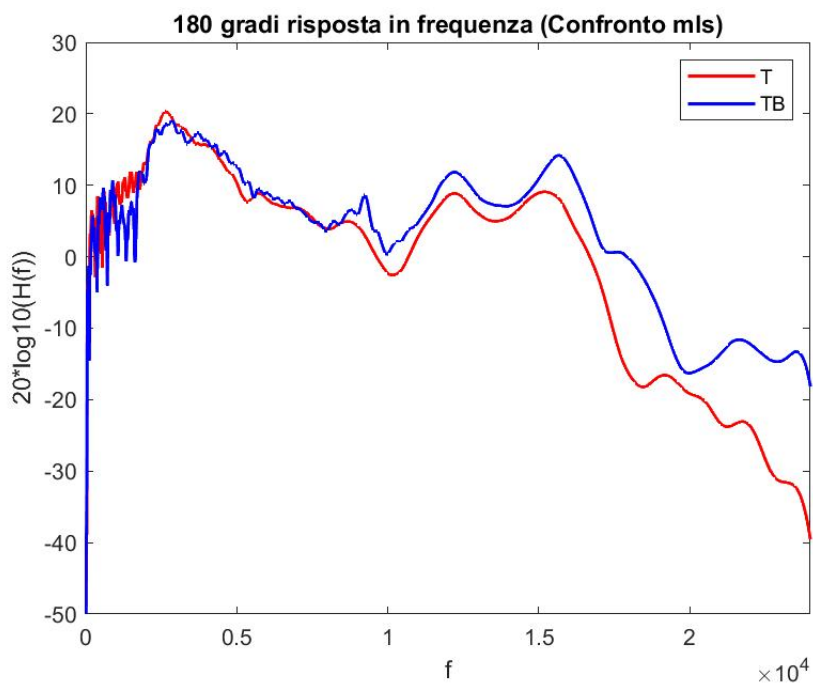


Figura 4.11: Confronto configurazioni (180° metodo mls)

4.6 Valutazione delle risposte tra diversi soggetti umani

Dopo aver analizzato le risposte ottenute utilizzando i manichini binaurali, sono state effettuate delle misure su quattro persone diverse (nel corso della trattazione verranno chiamati P1, P2, P3 e P4), tra cui due donne (P1 e P4) e due uomini (P2 e P3), con età comprese tra i 25 e 35 anni, utilizzando dei microfoni interaurali "Sennheiser mke 2-ew gold" posti all'ingresso del canale uditivo per registrare gli stimoli.



Figura 4.12: Posizionamento soggetto reale

4.6.1 Variazione dell'altezza della sorgente

Come nel caso visto precedentemente nella configurazione TB del manichino, la sorgente è stata elevata di 7 cm e di 31 cm rispetto all'altezza di partenza, ovvero quella delle orecchie del soggetto.

Partendo da **P1**, si vede l'andamento dell'orecchio sinistro ad un angolo di 100°: in Figura 4.13 è possibile vedere molto bene che, all'aumentare dell'altezza, il notch presente intorno ai 10 kHz si sposta leggermente a frequenze più basse, mantenendo

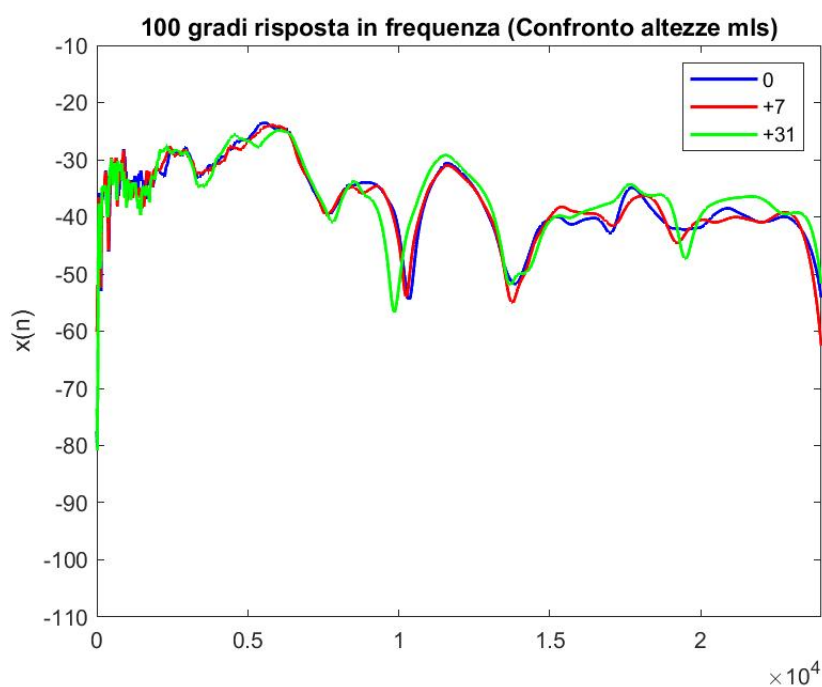


Figura 4.13: Confronto delle altezze del soggetto P1

un fattore di attenuazione di circa 55 dB. Inoltre, a frequenze più alte, in particolare dai 16 kHz in poi, sollevando la sorgente si vede che la larghezza del picco presente a 17.7 kHz a 0 cm aumenta fino a creare un notch intorno ai 19.5 kHz a 31 cm.

È stato considerato quindi l'andamento della risposta impulsiva dell'orecchio sinistro ad un angolo di 180° di **P2**: anche in questo caso, Figura 4.14, il notch si sposta sempre verso frequenze più basse aumentando l'altezza, ma questa volta dai 6.7 kHz ai 5.9 kHz con un fattore di attenuazione di circa 50 dB. In questo caso si verifica la trasformazione di un notch in un picco intorno ai 12 kHz, dove la differenza fra loro è di circa 30 dB.

In Figura 4.15 viene visualizzato l'andamento dell'orecchio destro ad un angolo di 260° di **P3**: in questo caso, non si verifica nessuno shift di picchi o notch elevando la sorgente, ma solamente l'eliminazione di due notch piuttosto larghi presenti a 6.2 kHz e 15.3 kHz.

Infine, seguendo l'andamento della risposta dell'orecchio sinistro ad un angolo di 100° di **P4**, l'unica differenza sostanziale ed evidente che si nota in Figura 4.16, è lo

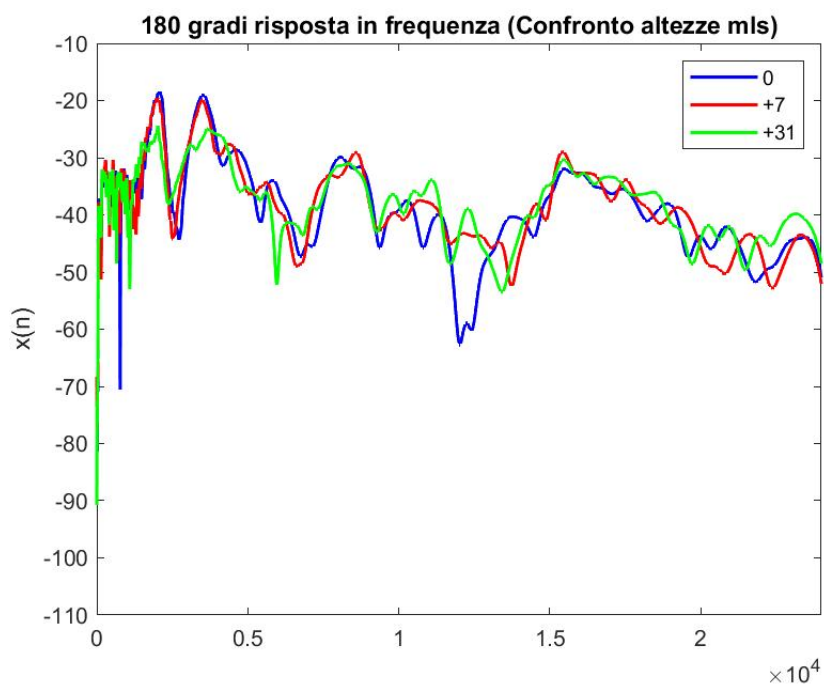


Figura 4.14: Confronto delle altezze del soggetto P2

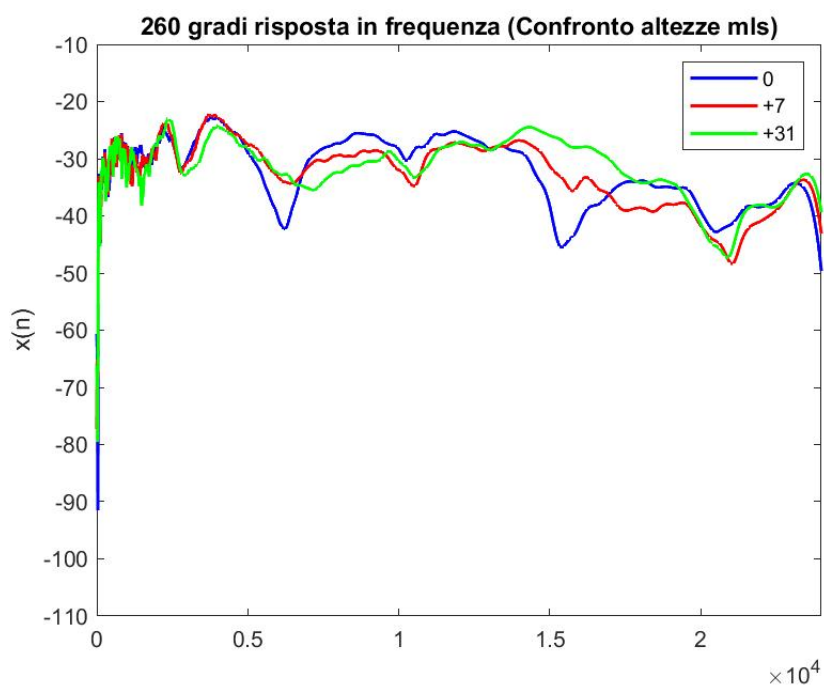


Figura 4.15: Confronto delle altezze del soggetto P3

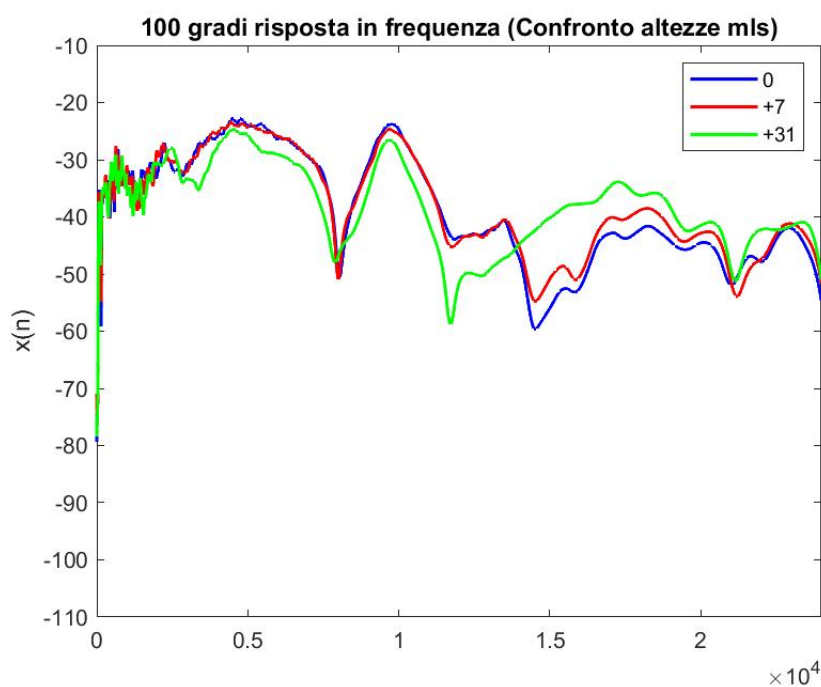


Figura 4.16: Confronto delle altezze del soggetto P4

spostamento alle basse frequenze del picco di attenuazione da 14.5 kHz a 11.7 kHz. Nella maggior parte dei soggetti si è visto che, aumentare l'altezza della sorgente può spostare picchi di attenuazione verso le basse frequenze oppure eliminare notch che erano presenti a 0 cm.

4.6.2 Analisi della simmetria tra l'orecchio destro e l'orecchio sinistro

In Figura 4.17 si vede la simmetria tra i padiglioni auricolari dei vari soggetti con un offset di 80°. In media, tutti i soggetti hanno una risposta in frequenza tra le due orecchie comparabili fino a circa 5 kHz, mentre a frequenze superiori gli andamenti divergono fra di loro, anche se in alcuni punti le risposte seguono un andamento simile, con qualche dB di differenza e una traslazione in frequenza. Rispetto al caso dei manichini binaurali, la differenza tra i due padiglioni è molto più grande.

Ad esempio, nel grafico (b) Figura 4.17 di P2, l'andamento dell'orecchio destro da 5 kHz a 6.5 kHz e l'andamento dell'orecchio sinistro da 6.7 kHz a 8.6 kHz è lo stesso ma si

trova a frequenze diverse. Altri andamenti simili possono essere visualizzati negli altri grafici. Questo non è solamente causato dalla diversa forma dei padiglioni auricolari, ma probabilmente anche dal posizionamento non identico dei microfoni interaurali all'ingresso del canale uditivo di ciascun orecchio che possono influire sull'accuratezza delle risposte.

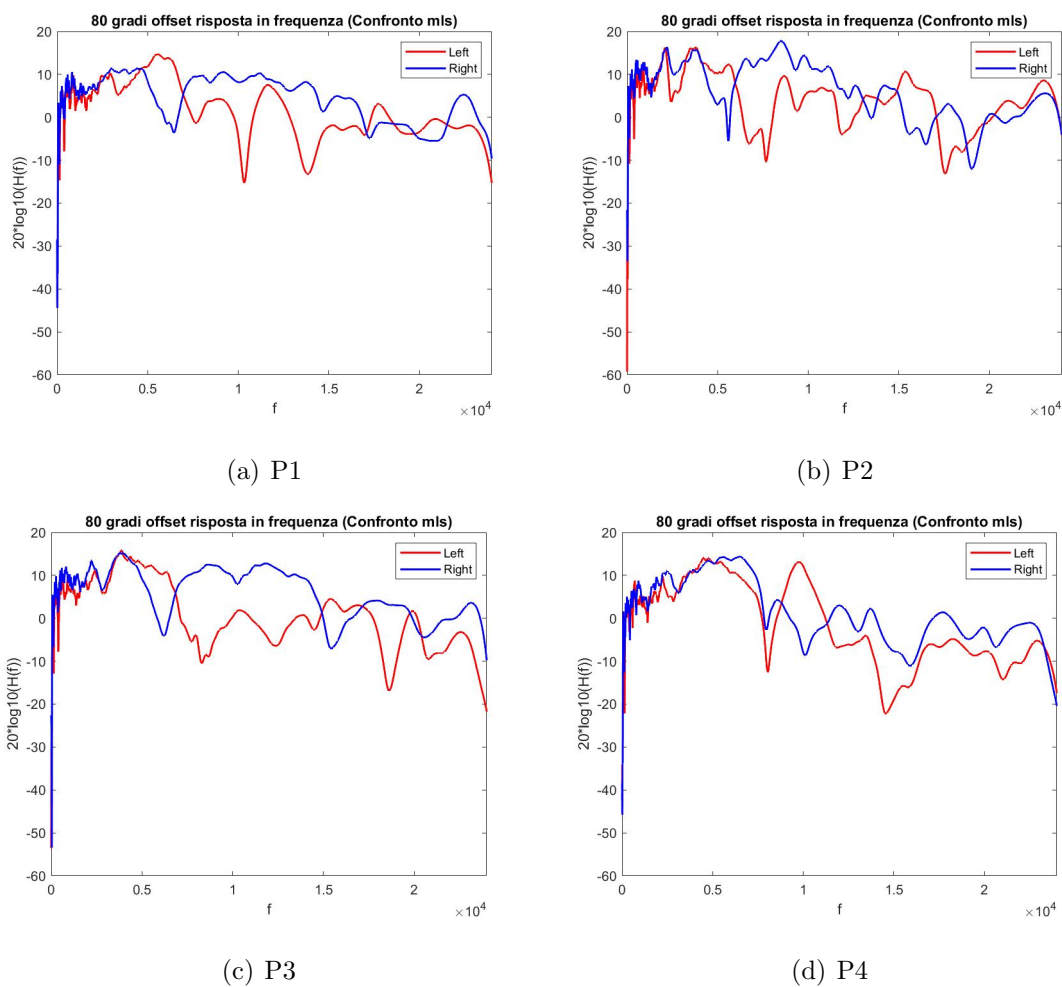


Figura 4.17: Simmetria soggetti (80° offset metodo MLS)

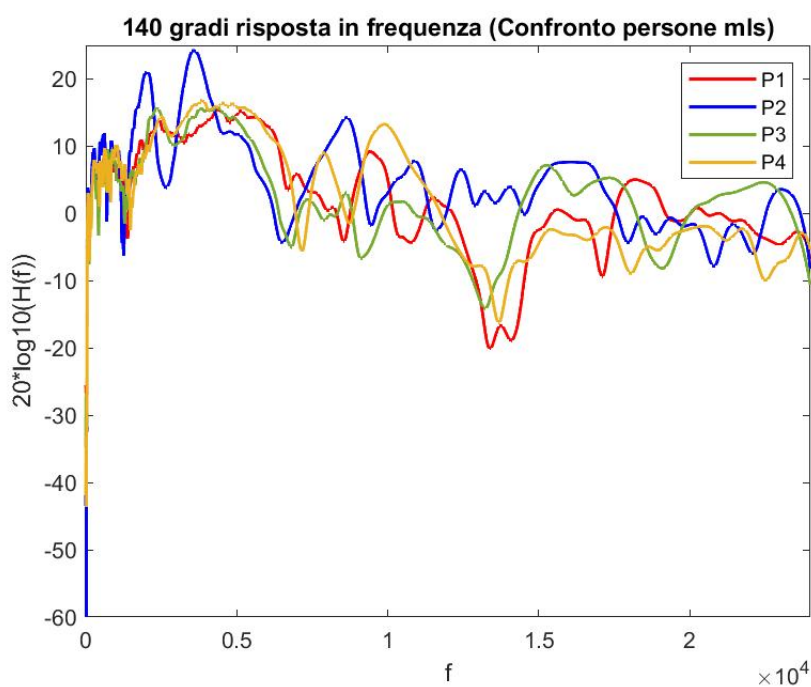
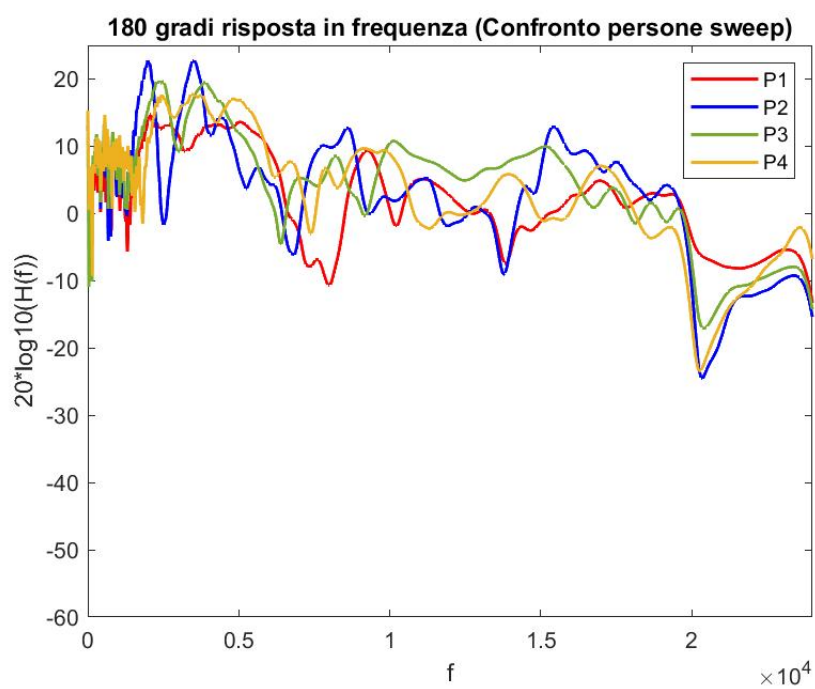
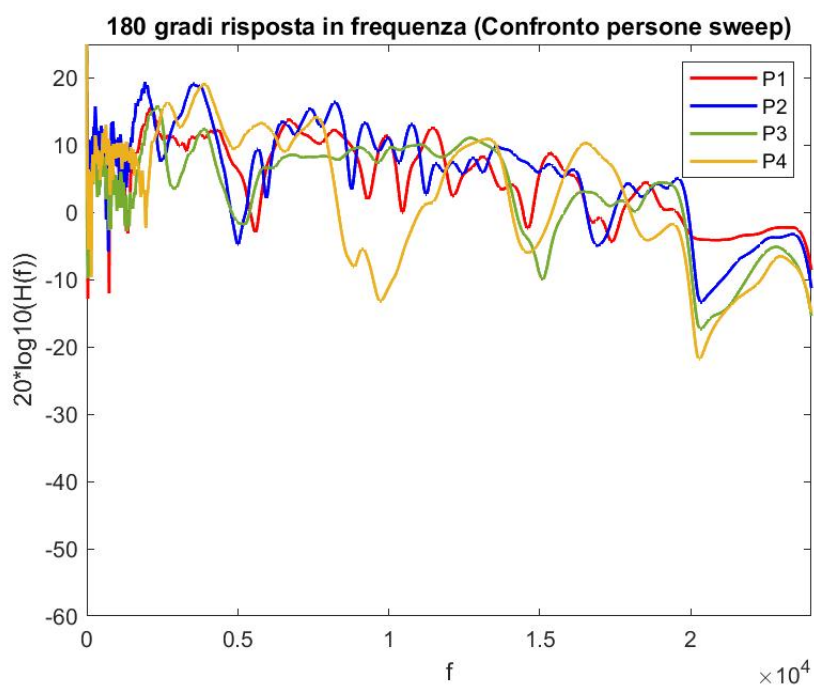


Figura 4.18: Risposte in frequenza dei 4 soggetti a 140° per l'orecchio sinistro

Se si sovrappongono le 4 risposte in frequenza, in Figura 4.18 ogni soggetto ha diversi andamenti, ma in alcuni punti, in questo caso intorno ai 6 kHz e ai 13.5 kHz, la maggior parte dei soggetti ha un picco di attenuazione. Queste somiglianze però, non si verificano in tutti gli angoli di misurazione, infatti, basti vedere l'andamento delle HRTF dei 4 soggetti nella direzione frontale sia per l'orecchio destro che per quello sinistro, Figura 4.19.



(a) Orecchio sinistro



(b) Orecchio destro

Figura 4.19: Risposte in frequenza dei 4 soggetti a 180°

Conclusioni

Nel corso delle analisi dei dati ottenuti, si è visto in modo molto evidente come ogni soggetto (manichino o soggetto reale) abbia delle HRTF con un andamento unico e che variano in modo complesso a seconda della posizione della sorgente sonora (la asimmetria dei padiglioni auricolari rende il tutto ancora più complesso), rendendo impossibile la possibilità di generare un ambiente virtuale tridimensionale con un unico set di HRTF che soddisfi ogni ascoltatore.

Durante l'acquisizione dei dati, è apparso che l'utilizzo del codice di Golay abbia portato via molto tempo, ritornando una TF distorta in alcune frequenze, rendendola inutilizzabile per l'analisi dei dati, quindi, durante l'acquisizione delle HRTF dei soggetti reali, non è stata più utilizzata, permettendo di velocizzare tutta la campagna di acquisizione. I metodi MLS e sweep invece, non hanno mai dato problemi significativi di distorsione.

Si può affermare che la variazione dell'altezza della sorgente sonora abbia diverse ripercussioni su ogni soggetto, da spostamenti a comparse o scomparse di picchi o notch in frequenza, ma in modo simile per entrambi i padiglioni auricolari, nonostante la loro asimmetria.

Durante l'analisi dei soggetti reali si è vista una notevole differenza tra la HRTF sinistra e destra, non solo a causa della asimmetria dei padiglioni auricolari, ma probabilmente anche a causa del posizionamento non identico dei microfoni interaurali all'ingresso del canale uditivo di ciascun orecchio, o il movimento della testa, seppur piccolo, può aver influito sulla precisione dei dati.

Sviluppi futuri potrebbero essere orientati verso la verifica di questi aspetti, realizzando una nuova campagna di acquisizione su uno dei manichini binaurali, posizionando

gli stessi microfoni interaurali utilizzati sui soggetti all'ingresso del canale uditivo del manichino, e fare poi un confronto tra le IR del manichino stesso e quelle dei microfoni interaurali.

Bibliografia

- [1] E. A. Macpherson and J. C. 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, no. 5, pp. 2219–2236, 2002.
- [2] B. Grothe, M. Pecka, and D. McAlpine, “Mechanisms of sound localization in mammals,” *Physiological reviews*, vol. 90, no. 3, pp. 983–1012, 2010.
- [3] Y. Zheng, J. Swanson, J. Koehnke, and J. Guan, “Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: A review,” *American Journal of Audiology*, pp. 1–16, 2022.
- [4] c. i. cheng and g. h. wakefield, “introduction to head-related transfer functions (hrtfs): representations of hrtfs in time, frequency, and space,” *journal of the audio engineering society*, september 1999.
- [5] D. W. Batteau, “The role of the pinna in human localization,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 168, no. 1011, pp. 158–180, 1967.
- [6] J. C. Middlebrooks and D. M. Green, “S () und localization by human listeners,” *Annu. Rev. Psychol*, vol. 42, pp. 135–59, 1991.
- [7] H. G. Fisher and S. J. Freedman, “The role of the pinna in auditory localization.” *Journal of Auditory research*, 1968.

- [8] C. Brown and R. Duda, “An efficient hrtf model for 3-d sound,” in *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997, pp. 4 pp.–.
- [9] E. M. 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, no. 1, pp. 111–123, 1993.
- [10] P. Bilinski, J. Ahrens, M. R. Thomas, I. J. Tashev, and J. C. Platt, “Hrtf magnitude synthesis via sparse representation of anthropometric features,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4468–4472.
- [11] S. Li and J. Peissig, “Measurement of head-related transfer functions: A review,” *Applied Sciences*, vol. 10, no. 14, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/14/5014>
- [12] S. Müller and P. Massarani, “Transfer-function measurement with sweeps,” *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.
- [13] C. Dunn and M. J. Hawksford, “Distortion immunity of mls-derived impulse response measurements,” *Journal of the Audio Engineering Society*, vol. 41, no. 5, pp. 314–335, 1993.
- [14] M. Golay, “Complementary series,” *IRE Transactions on Information Theory*, vol. 7, no. 2, pp. 82–87, 1961.
- [15] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, “Comparison of different impulse response measurement techniques,” *Journal of the Audio engineering society*, vol. 50, no. 4, pp. 249–262, 2002.