



**UNIVERSITA' DEL SALENTO**

---

FACOLTÀ DI INGEGNERIA

Corso di Laurea in Ingegneria dell'Informazione

TESI DI LAUREA IN  
SEGNALI E SISTEMI

## **Elaborazione del segnale audio 3D in MATLAB**

Laureando:  
**Salvatore Corvaglia**

Relatore:  
**Prof. Angelo Coluccia**

# Indice 1

## Prefazione

<i>Obiettivo e organizzazione della tesi. . . . .</i>	<i>4</i>
---	----------

## 1. Il suono nello spazio

<i>1.1 3D Audio System . . . . .</i>	<i>5</i>
<i>1.2 Il sistema di riferimento . . . . .</i>	<i>7</i>
<i>1.3 Percezione del suono nello spazio . . . . .</i>	<i>8</i>
<i>1.4 ITD - Interaural Time Difference . . . . .</i>	<i>10</i>
<i>1.5 ILD - Interaural Level Difference . . . . .</i>	<i>11</i>

## 2. Registrazione binaurale

<i>2.1 Tecniche di registrazione binaurale . . . . .</i>	<i>14</i>
<i>2.2 Riproduzione binaurale . . . . .</i>	<i>16</i>
<i>2.3 Stereofonia e binauralità . . . . .</i>	<i>18</i>
<i>2.4 Audio transaurale e binaurale . . . . .</i>	<i>18</i>

## 3. Modello matematico di HRIR e HRTF

<i>3.1 Definizioni di HRIR e HRTF. . . . .</i>	<i>22</i>
<i>3.2 Caratteristiche nel dominio del tempo e frequenza. . . .</i>	<i>23</i>

3.3	<i>Dati e strumenti di analisi . . . . .</i>	25
3.4	<i>Database di dominio pubblico . . . . .</i>	26
3.5	<i>Database KEMAR-MIT . . . . .</i>	27
3.6	<i>Implementazione del codice MATLAB . . . . .</i>	29
	 <b>Conclusioni</b>	 35
	 <b>Bibliografia</b>	 36
	 <b>Appendice</b>	 37
	 <b>Ringraziamenti</b>	 41

## Prefazione

Quando si ascolta un suono, spesso ci si domanda da dove provenga e si presta attenzione alle differenze tra ciò che si sente da un orecchio o dall'altro.

Girando la testa ci si rende conto che la percezione del suono cambia.

Queste variazioni non sono dovute solamente alla posizione della sorgente sonora, bensì alla forma di ogni singolo orecchio. I suoni avvertiti infatti sono elaborati e trasmessi al cervello attraverso un filtro naturale: l'organo uditivo.

Dal fatto che la forma dell'orecchio cambia da persona a persona, segue che la percezione del suono nello spazio è diversa per ogni individuo. Su questo scenario si pone il lavoro svolto durante la realizzazione di questa tesi. In particolare, oggetto di studio sono state le tecniche binaurali per l'ascolto del suono, tenendo conto delle varie trasformazioni che modificano l'onda sonora nel percorso da sorgente a soggetto ascoltatore.

Considerando l'insieme ambiente-uomo come un sistema LTI si può determinare la risposta all'impulso, detta HRIR (*Head Related Impulse Response*), e la sua relativa Trasformata di Fourier HRTF (*Head Related Transfer Function*).

## **Obiettivo e organizzazione della tesi**

L'obiettivo di questo lavoro di tesi è lo sviluppo di un modello di elaborazione del segnale audio 3D tramite MATLAB, un linguaggio di alto livello e un ambiente interattivo per il calcolo numerico, l'analisi, la programmazione e la visualizzazione dei dati.

Nei Capitoli 1 e 2 descriveremo la spazializzazione del suono, nel quale si espongono i modelli matematici più utilizzati per modellare, tramite tecniche binaurali, lo spostamento di una sorgente nello spazio.

Verranno esposti gli effetti causati dalle parti del corpo sul suono che giunge alle orecchie e, quindi, dei parametri principali (quali ITD ed ILD) e delle funzioni che descrivono i segnali binaurali in base alla posizione della sorgente rispetto alla testa.

Nel Capitolo 3 analizzeremo in modo dettagliato la risposta all'impulso HRIR, la sua trasformata HRTF e il progetto in MATLAB da cui trarremo le nostre conclusioni, commentando i risultati ottenuti.

# Capitolo 1

## Il suono nello spazio

In questo capitolo introduttivo tratteremo i concetti basilari dell'audio 3D, evidenziando il meccanismo attraverso il quale l'apparato uditivo umano percepisce un suono come localizzato in una ben determinata posizione e come l'onda sonora propagata nello spazio venga modificata dalle diverse parti del soggetto ricevitore.

In particolare si vedranno i concetti di ITD e ILD e i sistemi di riferimento per descrivere la posizione della sorgente rispetto all'ascoltatore.

### 1.1 3D Audio System

Per suono si intende quel fenomeno meccanico dato da una perturbazione di un mezzo di trasmissione (in genere l'aria) e che abbia caratteristiche tali da essere percepito dall'orecchio umano.

L'equazione che regola la propagazione del suono è quella di d'Alembert:

$$\nabla^2 p(\mathbf{x}, t) = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}(\mathbf{x}, t)$$

dove  $\mathbf{x}$  rappresenta il vettore delle coordinate Euclidee nello spazio,  $p$  la pressione acustica e  $c$  la velocità di propagazione del suono.

Le caratteristiche dell'ascolto umano sono gli elementi che si vogliono modellare per la realizzazione del suono 3D.

La progettazione e implementazione di ambienti virtuali può essere divisa in tre passi:

- 1) **definizione:** La definizione di un ambiente acustico virtuale include la conoscenza delle caratteristiche del sistema da realizzare, informazioni sulla sorgente sonora, sulla geometria dell'ambiente e sui soggetti ascoltatori.
- 2) **modellizzazione:** La parte di modellizzazione è divisa in tre parti che rispecchiano il paradigma sorgente - mezzo - ricevitore.
- 3) **riproduzione:** Metodi per ricreare l'illusione di un campo sonoro tridimensionale con l'aiuto di trasduttori e tecniche di signal processing.

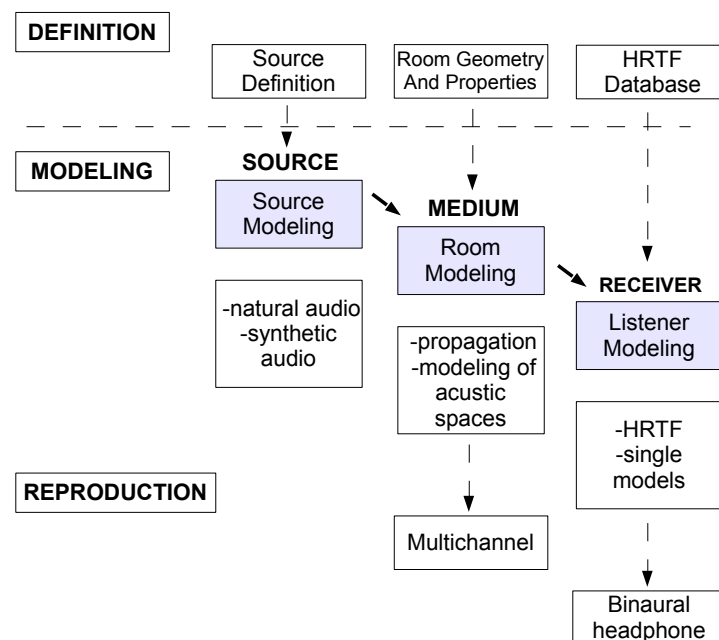


Figura 1: Concetti per la modellazione di un ambiente

## 1.2 Il sistema di riferimento

Le caratteristiche dell'ascolto umano sono gli elementi che si vogliono modellare per realizzare un suono 3D.

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

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

Il sistema di coordinate può essere del tipo *verticale-polare* o *interaurale-polare*. In entrambi i casi, le dimensioni che descrivono lo spazio tridimensionale sono:

- azimuth  $\theta$ , coordinata angolare orizzontale
- elevazione  $\phi$ , coordinata angolare verticale
- raggio  $r$ , coordinata radiale

Nel caso di sistema di coordinate verticale-polare (Figura 2a), l'azimut corrisponde all'angolo tra il piano mediano e il semipiano avente inizio sull'asse verticale  $z$  e contenente la sorgente sonora ( $\theta \in [-\pi; +\pi]$ ).

Sullo stesso semipiano, l'elevazione corrisponde all'angolo tra il piano orizzontale ed il segmento che congiunge l'origine degli assi e la sorgente sonora ( $\phi \in [-\pi/2; +\pi/2]$ )



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

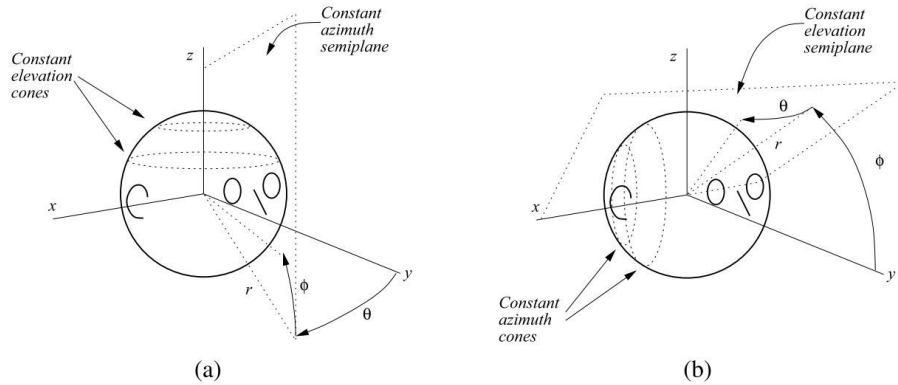


Figura 2: Sistema di coordinate verticale-polare (2a) e sistema di coordinate interaurale-polare (2b)

### 1.3 Percezione del suono nello spazio

Il suono prodotto da una sorgente è soggetto a diverse trasformazioni nel cammino verso le orecchie dell'ascoltatore, tra cui le perdite di energia isofrequenziali dovute alla distanza percorsa e quelle dipendenti dalla frequenza del suono, dovute quest'ultime agli oggetti schermanti che l'onda incontra durante la propagazione.

I segnali che arrivano alle due orecchie dell'ascoltatore saranno quindi tra loro diversi anche a causa della presenza della testa. Proprio in base a tale differenza dei flussi sonori, l'ascoltatore sarà in grado di discriminare direzione e distanza delle sorgenti, entro soglie di errore accettabili. Tale concetto è alla base della teoria sull'*audio binaurale*.

Infatti le parti anatomiche dell'uomo interferiscano con le onde sonore emesse da una sorgente acustica principalmente a causa dei fenomeni di riflessione e diffrazione.

Molte ricerche hanno dimostrato come l'uomo usi alcuni segnali di localizzazione.

I principali sono: la **ITD** (*Interaural Time Difference*) che è la differenza temporale tra gli istanti di arrivo di un suono alle due orecchie e si misura in secondi e la **ILD** (*Interaural Level Difference*), chiamata anche **IID** (*Interaural Intensity Difference*), che è il rapporto tra le ampiezze dei suoni in ingresso ai due canali uditivi, dovuto all'effetto schermante ad opera della testa. Esse variano in funzione dell'azimuth, dell'elevazione e della frequenza e forniscono le informazioni all'ascoltatore per localizzare un suono nello spazio.

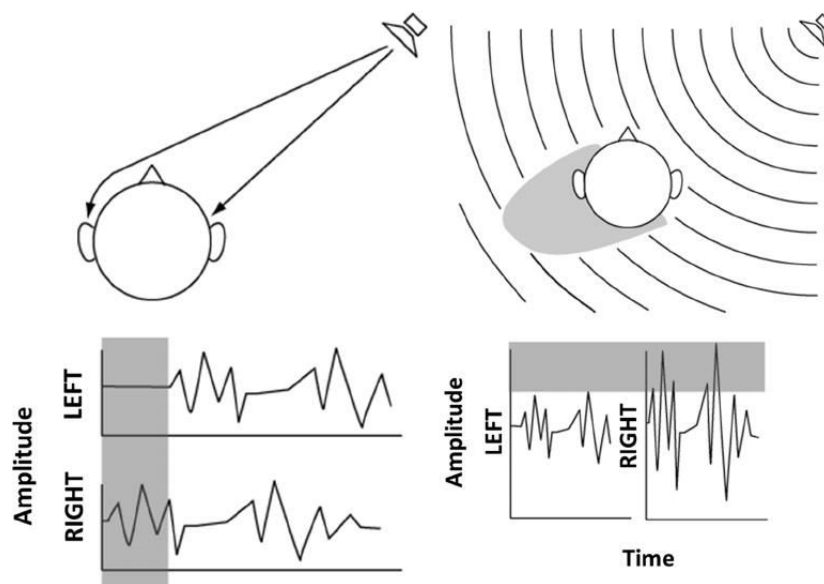


Figura 3: Rappresentazione schematica di ITD e ILD

## 1.4 ITD

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

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

Assumendo proprio l'ipotesi di testa sferica e sorgente sonora sufficientemente distante in modo da considerare planari le onde che raggiungono la testa, l'ITD può essere ricavata dividendo la distanza extra che l'onda deve percorrere per raggiungere l'orecchio più distante ( $\Delta x$ ) per la velocità del suono ( $c$ ).

Otteniamo:

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

dove  $a$  indica il raggio della testa,  $\theta$  l'angolo di azimut in radianti e  $c$  la velocità del suono in m/s.

Quindi la ITD è uguale a zero quando la sorgente è direttamente di fronte all'ascoltatore ( $\theta = 0$ ) ed ha il suo massimo ( $a/c(\frac{\pi}{2} + 1)$ ) quando la sorgente proviene dalla parte di una delle due orecchie ( $\pi = \theta/2$ ).

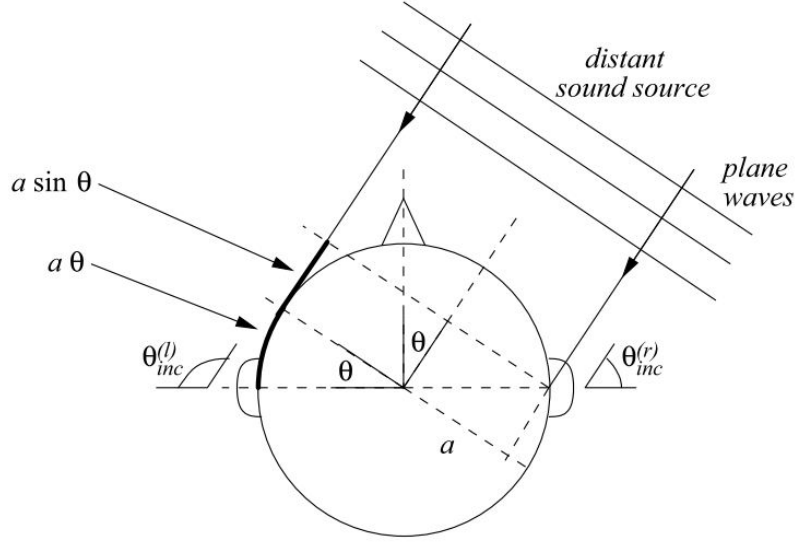


Figura 4: Stima dell'ITD in caso di sorgente sonora distante (onde piane) e testa sferica

## 1.5 ILD

Al contrario dell'ITD, l'ILD è fortemente legato alla frequenza: a basse frequenze alle due orecchie giunge una pressione sonora di differenza trascurabile l'una dall'altra, mentre, ad alte frequenze, la differenza diventa significativa.

Consideriamo la seguente situazione semplificata: testa sferica con raggio  $a$  e una sorgente sonora puntiforme posizionata alla distanza  $r > a$  dal centro della sfera.

La diffrazione di un'onda acustica sulla sfera vista da un punto scelto della sfera stessa è espressa dalla funzione di trasferimento:

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

dove  $\mu$  è la frequenza normalizzata e  $\rho$  la distanza normalizzata:

$$\mu = \frac{2\pi a}{c} f$$

$$\rho = \frac{r}{a}$$

e  $P_m$  e  $h_m$  sono rispettivamente il Polinomio di Legendre di ordine  $m$  e la funzione di Hankel sferica e  $\theta_{inc}$  è l'angolo di incidenza.

Dal grafico in Figura 5 si può notare che, a basse frequenze la funzione di trasferimento non è dipendente dalla direzione del suono e l'ampiezza è in prossimità dell'unità per ogni angolo di incidenza; ad alte frequenze ( $\mu > 1$ ) si fa sentire la dipendenza direzionale; nel caso di incidenza normale ( $\mu = 0$ ) si osserva un guadagno di 6 dB, in quanto questa situazione è paragonabile ad un'onda piana che incide perpendicolarmente con una superficie piana rigida.

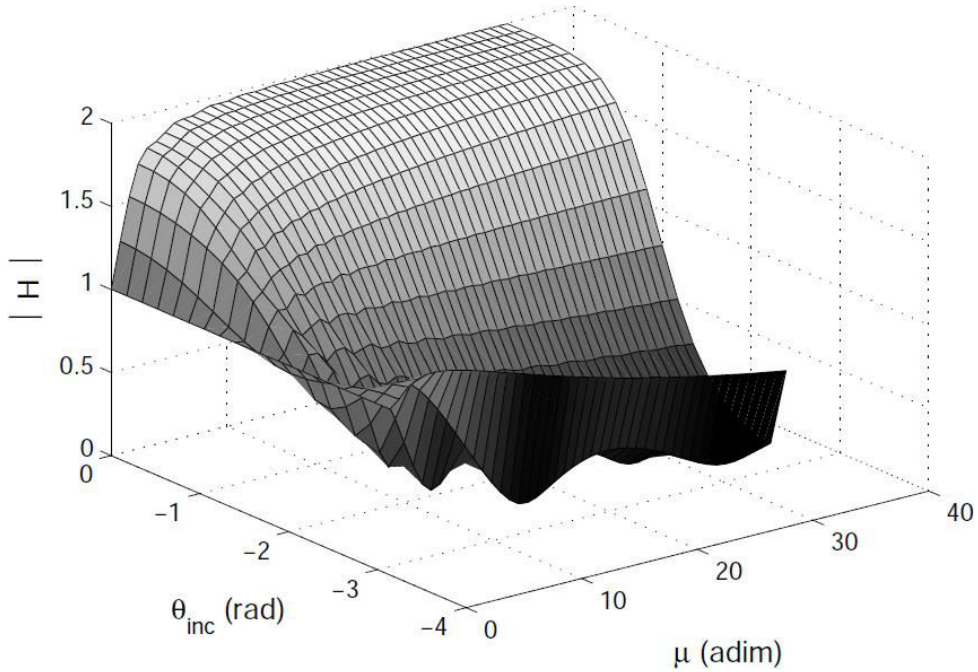


Figura 5: Grafico della funzione  $|H_{sfera}(\rho, \theta_{inc}, \mu)|$

La misura dell' ITD e dell'ILD non permette di localizzare la distanza della sorgente ma solo la direzione di provenienza del suono.

Il nostro sistema uditivo utilizza altre strategie per valutare la distanza della sorgente:

- in ambienti chiusi esso è in grado di valutare quanta dell'energia sonora catturata dai padiglioni auricolari arriva direttamente dalla sorgente e quanta da fenomeni di riflessione con le pareti: dalla proporzione di questi due contributi il sistema uditivo è in grado di stimare la distanza della sorgente.

In condizioni anecoiche, vale a dire in un ambiente un ambiente progettato per ridurre il più possibile la riflessione di segnali sulle pareti, la percezione della distanza di una sorgente è correlata all'*intensità* del segnale.

Infatti in queste condizioni, il guadagno di intensità sonora  $g$  è inversamente proporzionale alla distanza relativa  $d$  sorgente-ricevitore:

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

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

- in ambienti aperti, se la sorgente è lontana, si può assumere che i segnali direzionali siano indipendenti dalla distanza.

Avvicinando la sorgente vicino al soggetto, entrando così nel “*near field*” (circa 1,5m), la ITD resta quasi indipendente dalla distanza mentre l'ILD aumenta fortemente nell'intero spettro e in modo molto marcato alle basse frequenze.

## Capitolo 2

### Registrazione Binaurale

Come abbiamo già introdotto nel Capitolo 1, l'anatomia umana gioca un ruolo fondamentale nel modo in cui le onde sonore sono percepite, decodificate e analizzate dal cervello umano: la testa fa in modo che un suono giunto per primo all'orecchio sinistro impieghi alcuni  $\mu\text{s}$  prima di arrivare all'orecchio destro.

Inoltre le onde sonore saranno influenzate, nel loro tragitto, dalla forma della testa stessa e da altri elementi fisici andando così a creare variazioni specifiche per ogni ascoltatore.

Le registrazioni binaurali imitano questo sistema. E' un metodo di registrazione tridimensionale del suono che ha il fine di ottimizzare la registrazione per l'ascolto in cuffia, riproducendo le percezioni acustiche di un ascoltatore situato nell'ambiente originario di ripresa dell'evento sonoro, mantenendo le caratteristiche di direzione a  $360^\circ$ .

#### 2.1 Tecniche di registrazione binaurale

La tecnica di registrazione binaurale più semplice prevede l'utilizzo di due microfoni posti su un sostegno in direzione reciprocamente divergente, a circa 17 cm di distanza l'uno dall'altro.

Questo metodo non consente una registrazione binaurale vera e propria, in quanto non tiene conto dell'effetto fisico che la testa dell'ascoltatore ha sulla propagazione del suono.

Una tecnica più elaborata prevede l'utilizzo di una testa di manichino, costruita con dimensioni e materiali atti a riprodurre fedelmente l'assorbimento sonoro di una vera testa umana e soprattutto la sua funzione di separatore naturale tra i due canali uditivi. La testa riproduce con particolare fedeltà la forma dei padiglioni auricolari e i canali uditivi, all'interno dei quali vengono posti due microfoni ad alta fedeltà. In tal modo i microfoni captano il suono come risulta equalizzato e modificato in fase dalla testa, e quindi nel modo più simile a come l'avrebbe percepito un ascoltatore reale.

Esistono metodi ancora più sofisticati, che usano complessi apparati di equalizzazione, e metodi semplificati, tra cui quello che prevede l'utilizzo di un disco fonoassorbente (disco Jecklin) tra i due microfoni contrapposti, al posto della ben più complessa e costosa testa artificiale.



Figura 6: Testa artificiale utilizzata per registrazioni binaurali



Un altro metodo che sfrutta il principio della registrazione binaurale è l'*olofonia*, che prevede la registrazione tridimensionale del suono, con un'altissima fedeltà nella riproduzione che viene percepito nelle esatte coordinate spaziali nelle quali fu prodotto originariamente.

Il punto forte dell'olofonia sta nel fatto di illudere di avere la terza dimensione nella percezione del suono simulando le orecchie umane; questo è importante perché l'olofono, al fine di rappresentare al meglio le caratteristiche umane, assomiglia a una testa che ha due capsule microfoniche incorporate a simulare le orecchie.

I suoni sono catturati quindi attraverso una coppia di microfoni omnidirezionali separati proprio dal cranio al fine di aggiungere quel realismo a cui siamo quotidianamente abituati.

## **2.2 Riproduzione binaurale**

La registrazione binaurale viene riprodotta al meglio mediante l'ascolto in cuffia, in particolar modo tramite i modelli "*in-ear*"; l'ascolto tramite casse acustiche tende a confondere il senso di spazialità dovuto alla separazione dei canali apportata dalla testa artificiale.

Le tecnologie più sofisticate, tuttavia, disponendo da strumenti di equalizzazione psicodinamica e di teste artificiali di particolare precisione anatomica consentendo un eccellente ascolto anche tramite altoparlanti stereo.

Un'ulteriore complicazione è dovuta al fatto che fisiologicamente l'orecchio umano attenua fortemente il suono frontale nelle frequenze attorno a 5 kHz.

Dato che la registrazione binaurale con l'utilizzo di una testa artificiale

attenua naturalmente la frequenza predetta, in quanto si comporta esattamente come un orecchio vero e proprio, l'ascolto di tali registrazioni dovrebbe essere effettuato con una cuffia avente risposta in frequenza lineare, da ottenersi tramite apparati elettronici appositamente costruiti, o tramite un'equalizzazione da parte dello stadio di amplificazione.

La stereofonia appare quindi ideale per la registrazione in studio, quando gli strumenti suonano in ambienti insonorizzati, spesso in momenti diversi, e di conseguenza vi è la necessità di un ulteriore lavoro di post-produzione e di mixaggio.

La registrazione binaurale, invece, è ideale per ricreare la sensazione "live" di una sala da concerto, così come per ogni tipo di ripresa ambientale.

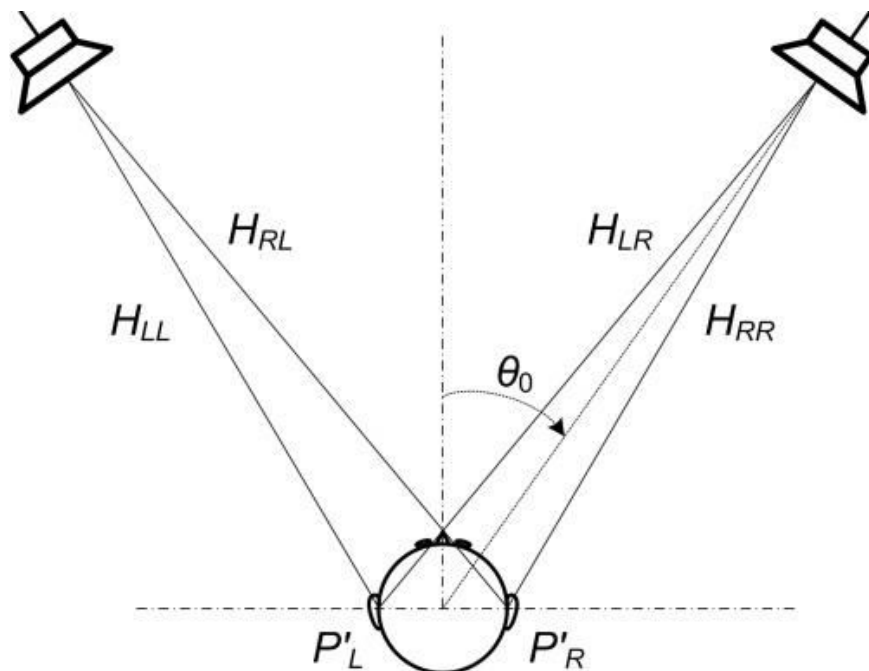


Figura 7: Riproduzione binaurale attraverso gli altoparlanti

## **2.3 Stereofonia e Binauralità**

La registrazione binaurale viene spesso confusa con quella stereofonica, ma in realtà si tratta di due sistemi molto differenti.

Nella registrazione stereofonica di un concerto, ad esempio, il suono viene tradizionalmente ripreso da numerosi microfoni di prossimità, strumento per strumento, e registrato su altrettante tracce separate, mentre l'acustica dell'ambiente di registrazione viene ripresa con una coppia di microfoni supplementari.

Il tutto dovrà poi essere miscelato su due sole tracce (destra e sinistra), per consentirne la comune riproduzione acusmatica.

In tal modo, nonostante vi sia la possibilità di intervenire sul suono di ogni strumento in fase di post-produzione, gli equilibri naturali tra i suoni degli strumenti e l'ambiente acustico della sala risulteranno compromessi o del tutto perduti, così che il suono registrato sarà inevitabilmente diverso da quello ascoltato direttamente al concerto.

Nella registrazione binaurale, invece, il suono viene ripreso complessivamente dal microfono a testa artificiale e registrato direttamente su due soli canali, garantendo una elevata somiglianza tra il suono ascoltato al concerto dal vivo e quello registrato.

## **2.4 Audio transaurale e binaurale**

Un'importante classificazione si basa sul metodo utilizzato per la riproduzione del materiale audio. Si distingue l'audio binaurale che fa uso di cuffie e l'audio transaurale che impiega un sistema di altoparlanti a due

canali. Entrambe presentano vantaggi e svantaggi che le fanno preferire per un'applicazione piuttosto che un'altra.

E' innanzitutto evidente, come con l'uso delle cuffie, l'orecchio destro sia interessato solo dal materiale audio del canale destro, e viceversa il sinistro. Questo permette un assoluto controllo da parte del progettista del sistema audio 3D sul materiale audio ascoltato evitando fenomeni di cross-talk che sono invece presenti in una riproduzione transaurale.

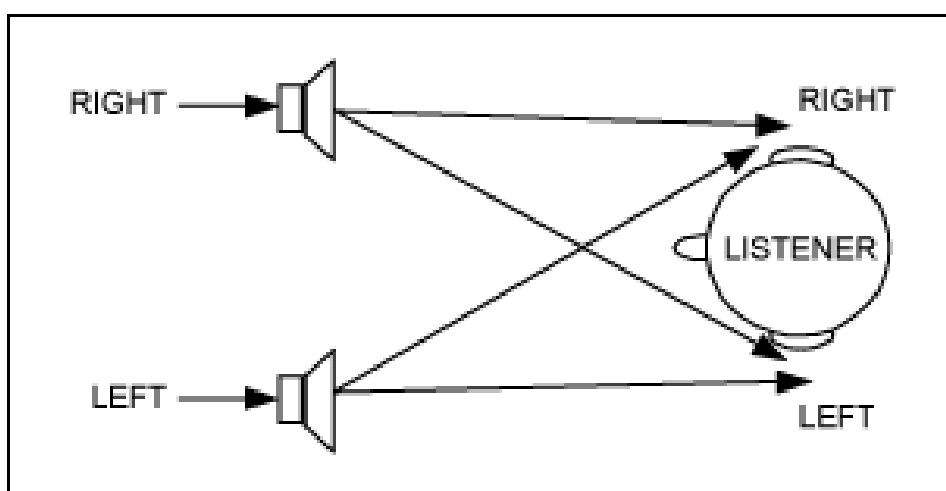


Figura 8: Fenomeni di crosstalk nell'ascolto transaurale

La libertà di movimento associata all'ascolto transaurale può però compromettere la resa del sistema audio 3D: se si considera l'ITD, risulta evidente come un movimento dell'ascoltatore verso un altoparlante costituisca una distorsione non prevedibile dal progettista e che quindi falserà la corretta collocazione della sorgente.

Viene così a individuarsi una regione spaziale limitata, detta sweet-spot (punto dolce), all'interno della quale le informazioni spaziali riprodotte saranno correttamente interpretate dal soggetto.

Al di fuori di essa, invece, la resa del sistema non è prevedibile ma in generale non porterà alla corretta spazializzazione.

E' invece ragionevole considerare che l'ITD nella riproduzione tramite cuffie sia minimo, visto che entrambi i segnali percorrono la stessa distanza.

## Capitolo 3

### Modello matematico di HRIR e HRTF

La propagazione delle onde sonore attraverso l'aria è un processo lineare e permette una semplice caratterizzazione matematica dei segnali sonori di localizzazione.

Quando la sorgente sonora e l'ascoltatore sono fissi, la trasmissione acustica da una sorgente puntiforme alle due orecchie può essere considerata come un sistema lineare tempo-invariante (LTI). Un *sistema dinamico lineare stazionario*, anche detto *sistema lineare tempo-invariante* o *sistema LTI*, è soggetto cioè al principio di sovrapposizione degli effetti e tale che il suo comportamento sia costante nel tempo.

Le funzioni di trasferimento riferite alla testa (HRTF) sono definite come le funzioni di trasferimento acustiche di questo sistema LTI:

$$H_L(r, \theta, \phi, f, a) = \frac{P_L(r, \theta, \phi, f, a)}{P_o(r, f)}, \quad H_R(r, \theta, \phi, f, a) = \frac{P_R(r, \theta, \phi, f, a)}{P_o(r, f)}$$

dove  $P_L$  e  $P_R$  rappresentano i livelli di pressione sonora all'orecchio, rispettivamente a destra e sinistra,  $P_o$  rappresenta il campo libero di pressione sonora al centro della testa,  $(r, \theta, \phi)$  le relative coordinate sferiche,  $f$  la frequenza e  $a$  l'individuo.

### 3.1 Definizioni di HRIR e HRTF

La pressione sonora prodotta da una sorgente nel canale uditivo è determinata esclusivamente dalla risposta impulsiva nel tragitto dell'onda sonora, chiamata **Head Related Impulse Response (HRIR)**.

Per ogni soggetto esistono diverse HRIR, le quali differiscono le une dalle altre per la posizione della sorgente sonora. Tale posizione è composta dall'azimuth, dall'elevazione e dalla distanza e discrimina appunto in modo univoco le HRIR di uno stesso soggetto. Nel dominio del tempo indichiamo le risposte all'impulso dell'orecchio sinistro e destro rispettivamente come  $h_L(t)$  e  $h_R(t)$ .

Definiamo  $x_L(t)$  e  $x_R(t)$  i segnali ricevuti dall'orecchio sinistro e destro ottenuti dalla seguente convoluzione:

$$x_{L,R}(t) = h_{L,R}(t) * x(t) = \int_{-\infty}^{\infty} h_{L,R}(t - \tau) x(\tau) d\tau$$

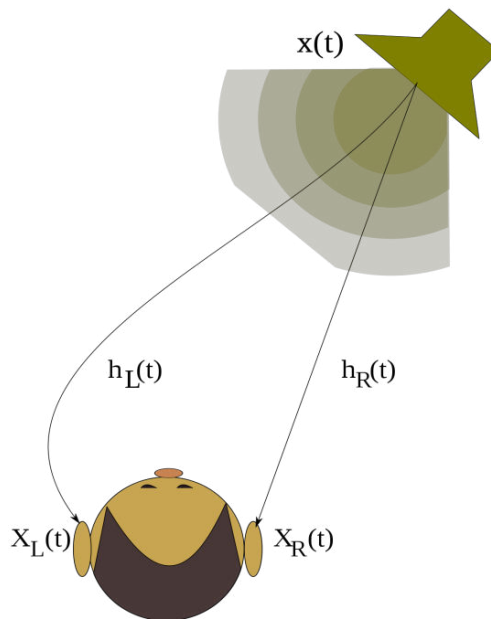


Figura 9: Filtraggio di un segnale  $x(t)$  di  $h_L(t)$  e  $h_R(t)$

Il suo corrispondente nel dominio della frequenza, ricavato attraverso la trasformata di Fourier, è la **Head Related Transfer Function (HRTF)**:

$$X_{L,R}(\omega) = F(h_{L,R}(t) * x(t)) = H_{L,R}(\omega)X(\omega)$$

In generale, tale funzione di trasferimento si indica con  $H^{(L),(R)}(r, \theta, \phi, \omega)$  (l'apice indica l'orecchio preso in considerazione). Nella maggior parte dei casi, si lavora per sorgenti *campo lontano* ed è quindi lecito svincolare l'HRTF dalla dipendenza da  $r$  stesso e scrivere  $H^{(L),(R)}(\theta, \phi, \omega)$ .

Si definisce quindi la HRTF ad un orecchio come il rapporto, dipendente da  $f$ , tra il livello di pressione del suono sul timpano ed il corrispondente livello di pressione al centro della testa in campo libero, ovvero come se l'ascoltatore fosse assente. Da cui:

$$H^{(L)}(\theta, \phi, \omega) = \frac{P^{(L)}(\theta, \phi, \omega)}{P_o(\omega)}, \quad H^{(R)}(\theta, \phi, \omega) = \frac{P^{(R)}(\theta, \phi, \omega)}{P_o(\omega)}$$

A causa della diversa conformazione anatomica, l'HRTF è una funzione propria di ogni ricevitore. Se ne può considerare una “media” che approssimi quella individuale con buona precisione. Quindi a seguito dell'operazione di trasformata, le HRTF sono convertite in decibel secondo la legge:

$$H_{dB} = 20 \log_{10}(H)$$

### 3.2 Caratteristiche nel dominio del tempo e frequenza

Anche se le HRIR o le HRTF variano tra individuo e individuo, si osservano alcune caratteristiche comuni in termini di tempo e frequenza.



Per facilitare la misurazione delle HRIR spesso viene utilizzato il **KEMAR** (*Knowles Electronic Manikin for Auditory Research*), strumento di ricerca acustica che permette misure riproducibili per la determinazione delle prestazioni degli apparecchi acustici e altri dispositivi elettroacustici.

La figura 10 mostra le HRIR di un KEMAR in campo lontano con azimuth a  $30^\circ$  e  $90^\circ$ . All'azimuth  $30^\circ$ , la grandezza HRIR, a 58 campioni è approssimativamente zero, corrispondente al ritardo di propagazione dalla sorgente sonora.

In pratica, una finestra di tempo è di solito applicata alle prime HRIR, e quindi il ritardo iniziale ha solo importanza relativa.

Il corpo principale delle HRIR, che riflette le complicate interazioni tra onde sonore incidenti e strutture anatomiche, persiste per circa 50-60 campioni.

Successivamente, la grandezza HRIR ritorna quasi a zero.

Quando la sorgente sonora si discosta da direzioni direttamente anteriori e posteriori, la differenza di ritardo iniziale nelle HRIR dell'orecchio destro e sinistro riflette la differenza di tempo di propagazione dalla sorgente, cioè l'ITD.

Ad azimuth  $90^\circ$ , per esempio, l'HRIR dell'orecchio sinistro ritarda all'HRIR dell'orecchio destro con un ritardo relativo di 28 campioni. Inoltre, quando la sorgente sonora si trova dal lato opposto all'orecchio interessato, ad esempio ad un azimuth di  $90^\circ$  per l'orecchio sinistro, la grandezza HRIR è visibilmente attenuata a causa dell'effetto testa ombra.

Come l'elevazione cambia, la differenza di ritardo iniziale e dell'entità tra le HRTF si riduce.

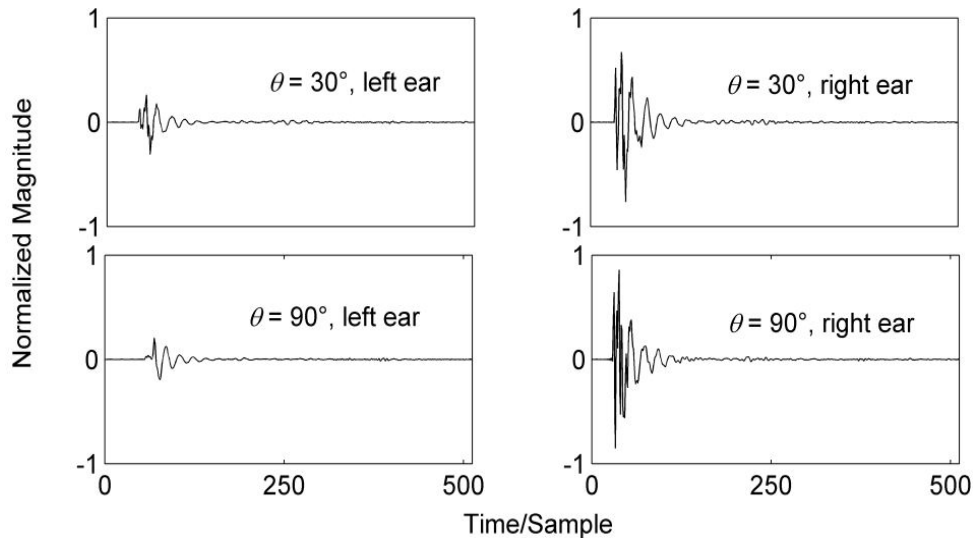


Figura 10: HRIR di un KEMAR in campo lontano con azimuth a  $30^\circ$  e  $90^\circ$

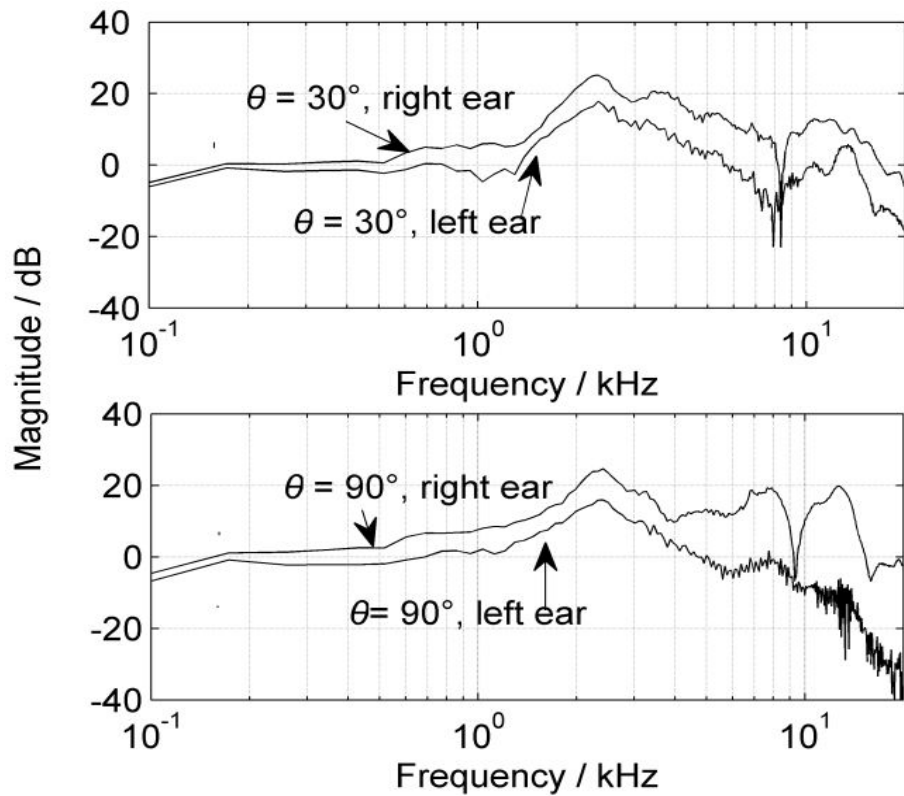


Figura 11: Ampiezze delle HRTF corrispondenti alle HRIR in Figura 10

### 3.3 Dati e strumenti di analisi

L'ascoltatore, ovvero la testa di un manichino, è posizionato al centro della sfera con un microfono ad entrambi gli orecchi.

Le risposte impulsive sono misurate eseguendo un segnale analogico e registrando le risposte alle orecchie per ogni posizione virtuale desiderata.

Ascoltatore ed altoparlanti non hanno bisogno di essere spostati, facilitando di molto le misurazioni. Tuttavia il posizionamento del microfono costituisce un problema: può essere applicato all'ingresso del canale uditivo. In molte applicazioni di 3D sound tipicamente si utilizza una singola HRTF per ogni utente. Un altro approccio sta nel personalizzare queste funzioni in base ad alcune caratteristiche o, alternativamente, di costruire matematicamente delle HRTF in base a caratteristiche comuni ad un certo numero di individui. Un'alternativa molto utilizzata sta nell'utilizzare le cosiddette “dummy heads” (teste di manichino), costruite mediando un consistente numero di misure. Il modello di manichino più utilizzato è proprio il KEMAR e i microfoni posizionati all'interno della dummy head hanno una risposta alle basse frequenze migliore dei microfoni sonda. L'uso del KEMAR ha sicuramente il vantaggio di rendere facilmente riproducibili i risultati e di evitare lo scomodo posizionamento dei microfoni nelle orecchie.

### **3.4 Database di dominio pubblico**

Ogni database di HRTF di dominio pubblico è stato realizzato secondo standard diversi (per ciò che riguarda gli incrementi di azimuth ed elevazione, le aree spaziali coperte, la durata e la frequenza di campionamento, i segnali di stimolo). Il MARL (*Music and Audio Research Laboratory*) della New York University ha definito un formato di file per la

memorizzazione di HRIR, verso il quale ha già convertito i dati relativi ai progetti di alcuni istituti di ricerca. Questi progetti sono chiamati:

-LISTEN

-CIPIC

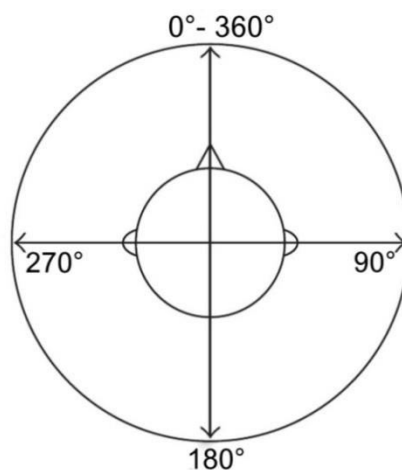
-FIU

-KEMAR-MIT

Quest'ultimo sarà usato nel nostro progetto MATLAB finale.

### 3.5 Database KEMAR-MIT

Bill Gardner e Keith Martin del MIT (Massachusetts Institute of Technology) hanno utilizzato proprio un KEMAR collocato in una camera anecoica e montato su una piattaforma girevole motorizzata. Il sistema di riferimento è ancora una volta quello verticale-polare, con l'altoparlante collocato a 1,4 m di distanza dal manichino. Nel complesso sono state effettuate 710 registrazioni per soggetto; 14 sono le elevazioni differenti, da  $-40^\circ$  a  $90^\circ$  con un incremento verticale di  $10^\circ$ , mentre il numero di posizioni azimuthali corrispondenti varia da 72 (incrementi orizzontali di  $5^\circ$ ) a solo 1 (per l'elevazione di  $90^\circ$ ).



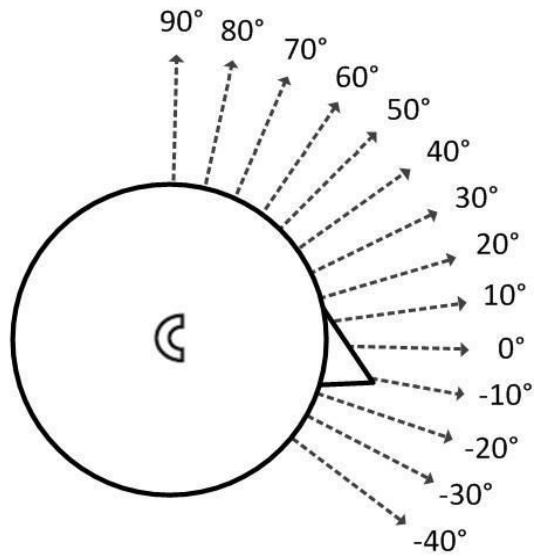


Figura 12 e 13: Angoli di azimuth ed elevazione

Elevazione	Numero di misurazioni	Incremento di Azimuth
-40	56	6.43
-30	60	6.00
-20	72	5.00
-10	72	5.00
0	72	5.00
10	72	5.00
20	72	5.00
30	60	6.00
40	56	6.43
50	45	8.00
60	36	10.00
70	24	15.00
80	12	30.00
90	1	x.xx

Tabella 1: Numero di misurazioni e incremento dell'angolo ad ogni elevazione

Il dataset di questo modello KEMAR-MIT è disponibile online al seguente

indirizzo: <http://sound.media.mit.edu/resources/KEMAR.html>

### 3.6 Implementazione del codice MATLAB

In questa sezione viene presentato l'esperimento condotto tramite MATLAB (*Matrix Laboratory*) della MathWorks, un linguaggio di alto livello e un ambiente interattivo per il calcolo numerico, la programmazione e la visualizzazione di dati, consentendo la creazione di modelli e applicazioni e lo sviluppo di algoritmi.

- La cartella *HRIRData* contiene i dati registrati delle singole HRIR ad elevazioni che vanno in intervalli di  $10^\circ$  che vanno da  $-40^\circ$  a  $90^\circ$ .
- Nello script *hrtfmodel.m* la prima parte del codice specifica il file di dati chiamati a seconda delle variabili di azimuth ed elevazione. Dopo aver specificato il file di dati, la funzione *read* legge quindi i dati HRIR richiesti dal file; viene utilizzata la funzione *fopen*, seguita da *fread* e *fclose* per aprire, leggere i dati dal file, e chiuderli nuovamente.

Nella seconda parte del codice, si utilizza la funzione *filter* per filtrare la variabile di ingresso attraverso la risposta all'impulso. Il segnale risultante viene poi diviso per il valore massimo assoluto di se stesso per evitare possibili distorsioni della forma d'onda. Il processo è fatto ovviamente per entrambi i canali destro e sinistro del segnale finale e viene riassunto in una matrice stereo output finale.

Nella parte finale del codice la funzione *wavwrite* produce un file wav di output con frequenza di campionamento  $44100\text{ Hz}$  a 32 bit. Le funzioni *strcat* e *num2str* vengono quindi utilizzate nuovamente, ma questa volta per etichettare il file di output basato sull'originale *Azi* ed *Ele* valori.

- Nello script *Sample\_Stereo\_Movement.m* si adopera la funzione *audioread* per leggere i dati del file audio d'ingresso '*corvaglia.wav*'.

Formato	: Wave
Dimensione	: 774 KiB
Durata	: 8s 986 ms
Modo bitrate generale	: Costante
Bitrate totale	: 706 kb/s
Audio	
Formato	: PCM
Impostazioni formato, Endianness	: Little
Impostazioni formato, Sign	: Signed
ID codec	: 1
Durata	: 8s 986 ms
Modalità bitrate	: Costante
Bitrate	: 705,6 kb/s
Canali	: 1 canale
Frequenza campionamento	: 44,1 kHz
Profondità bit	: 16 bit
Dimensione della traccia	: 774 KiB (100%)

Tabella 2: Caratteristiche principali del file audio *corvaglia.wav*

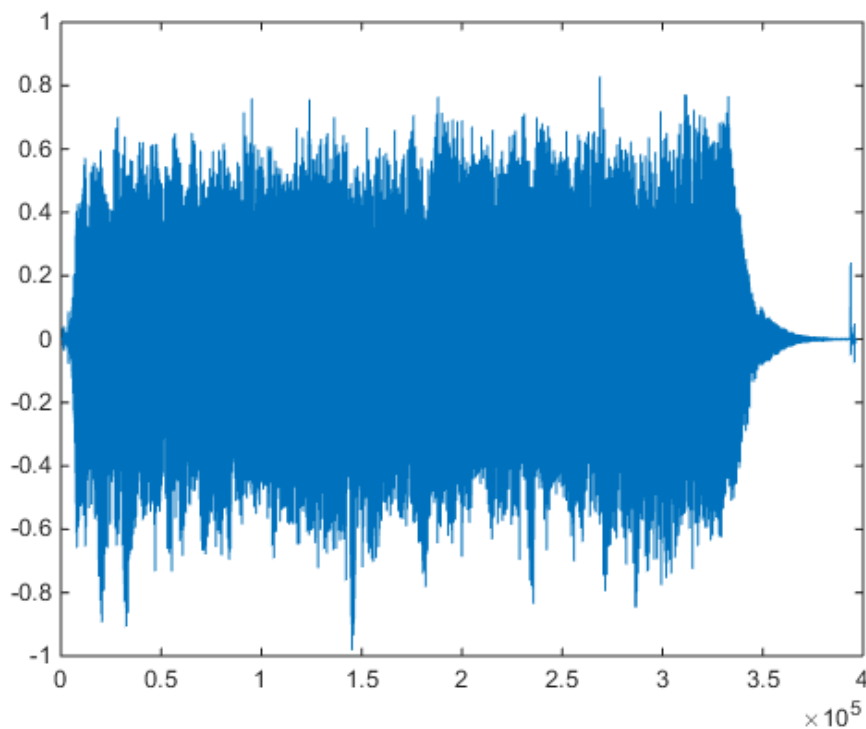


Figura 14: Grafico del segnale audio d'ingresso *corvaglia.wav*

Si utilizza una elevazione di  $10^\circ$  e angoli di azimuth a  $45^\circ$ ,  $120^\circ$ ,  $270^\circ$ ,  $320^\circ$  e con la funzione *vertcat* si concatenano le matrici verticalmente proprio per unire l'audio a diverse angolazioni.

Infine la funzione *sound* invia il segnale audio *s* alla frequenza di campionamento *fs* e la funzione *disp* mostra la relativa matrice delle *s*.

Nell'ordine, si riportano i grafici con le diverse variazioni di azimuth prese in considerazione e il grafico della concatenazione verticale delle HRTF in uscita.

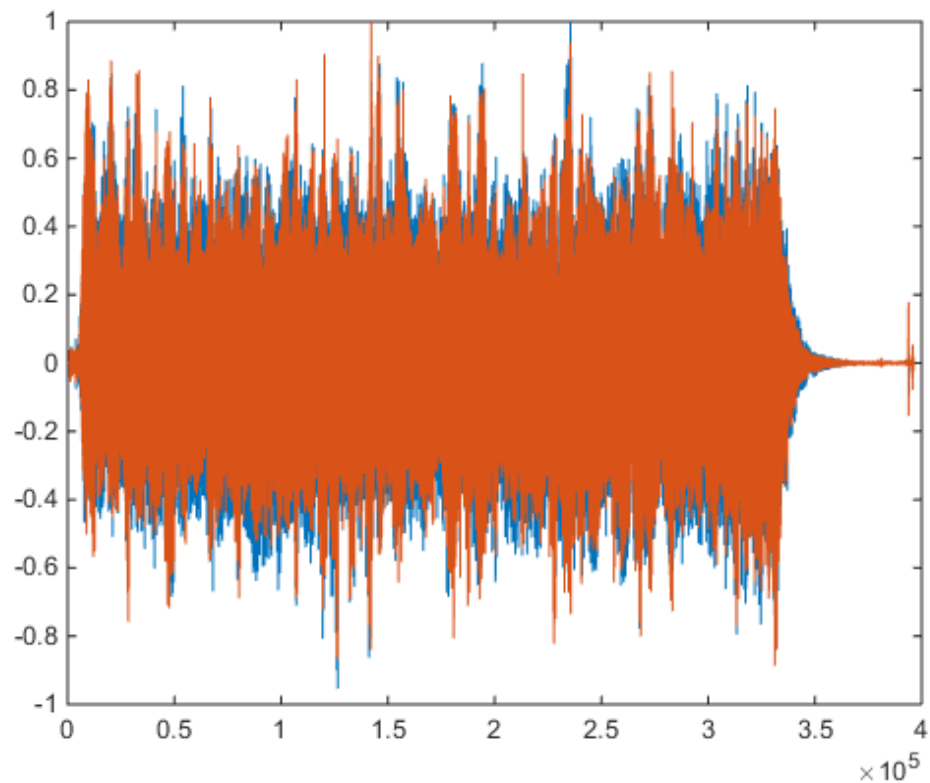


Figura 15: HRTF in uscita con Elevazione  $10^\circ$  e Azimut  $45^\circ$



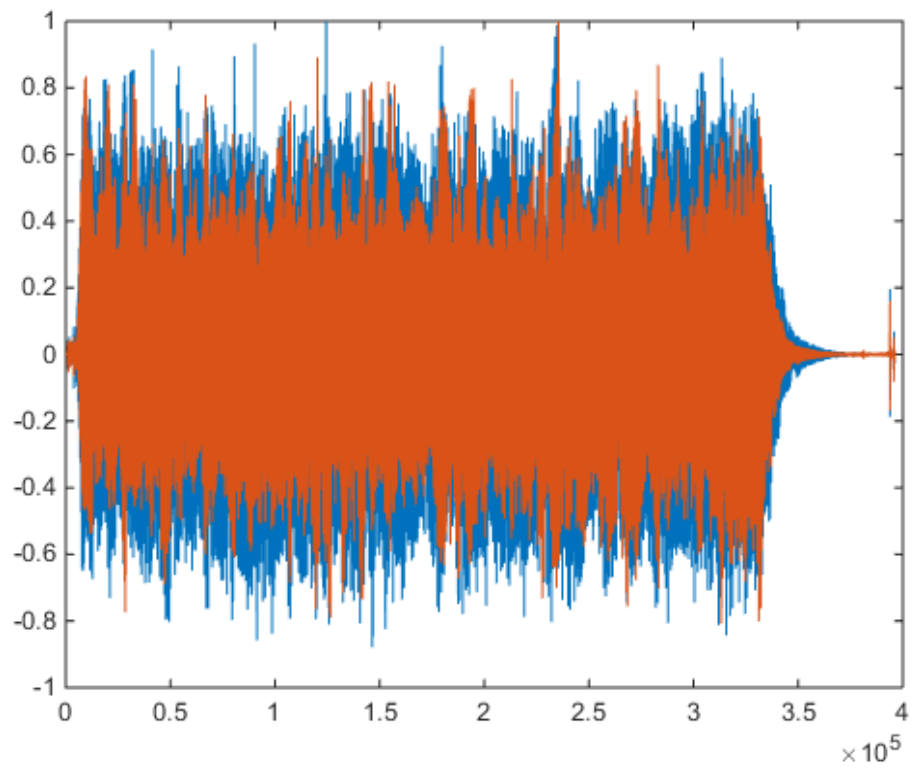


Figura 16: HRTF in uscita con Elevazione 10° e Azimut 120°

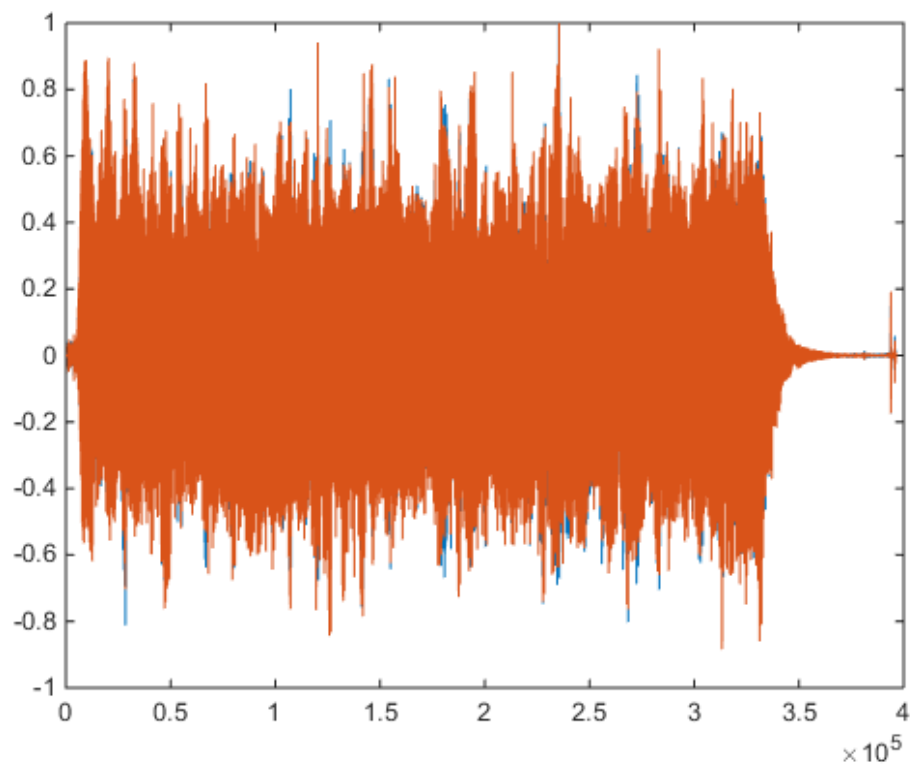


Figura 17: HRTF in uscita con Elevazione 10° e Azimut 270°

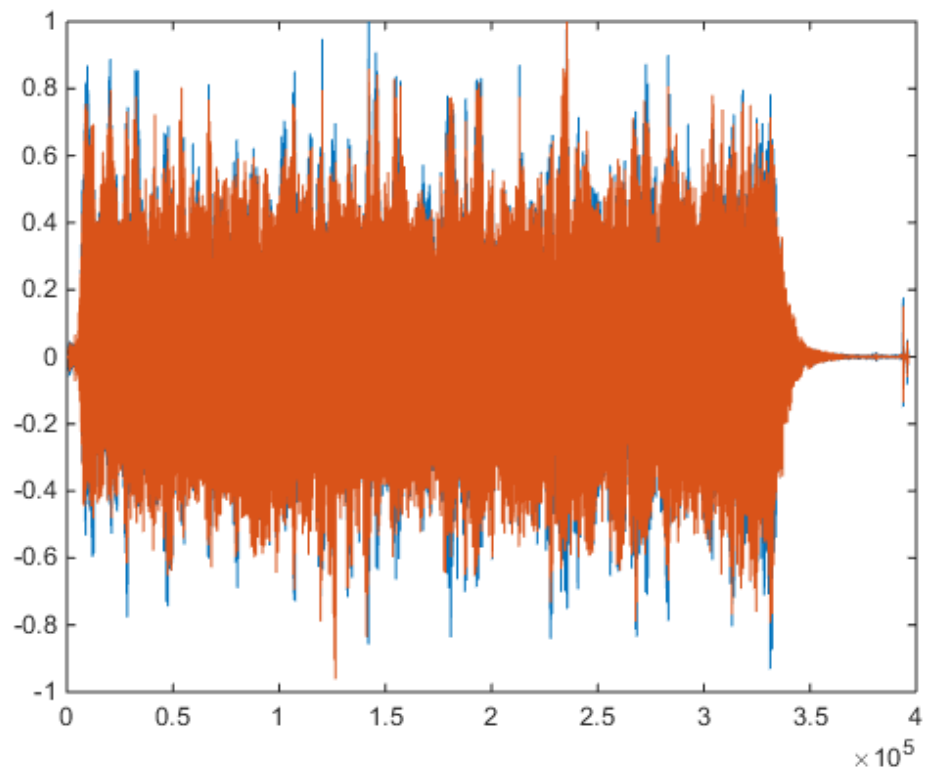


Figura 18: HRTF in uscita con Elevazione 10° e Azimut 320°

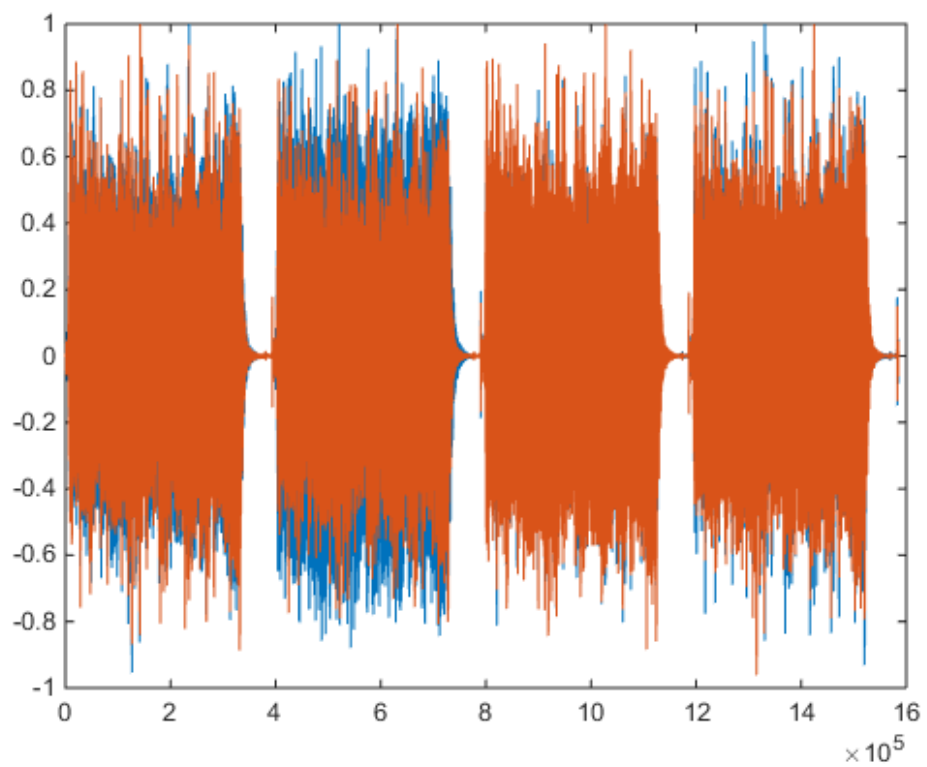


Figura 19: HRTF in uscita concatenate tra di loro con le diverse angolazioni

Nel caso si vuole registrare direttamente in MATLAB un segnale audio wav tramite agente esterno, come un semplice microfono, si fa uso nello script `Sample_Stereo_Movement.m` della funzione *audiorecorder* che va a creare un file audio con frequenza di campionamento 8000 Hz, 8 bit e 1 canale.

```
recObj = audiorecorder(8000, 16, 2); % Crea un file audio
get(recObj)

% Registra la tua voce per 5 secondi

recObj = audiorecorder;

disp('Start speaking.')

recordblocking(recObj, 5);

disp('End of Recording.');
```

```
myRecording = getaudiodata(recObj); % Memorizza i dati

filename = 'myRecording.wav';

audiowrite(filename, myRecording, 8000);
```

Successivamente svolge le stesse funzioni già descritte in precedenza producendo, oltre la registrazione effettuata, le relative HRTF in uscita.

# Conclusioni

Il lavoro di tesi si è concentrato sullo studio ed implementazione MATLAB di un algoritmo per la spazializzazione del suono tramite tecniche binaurali, tendendo in considerazione delle caratteristiche dell'ascoltatore.

È stato descritto un modello di HRTF personalizzate che possono essere usate in un ambiente per il rendering spaziale del suono in tempo reale.

È stato utilizzato un database di dominio pubblico, reso disponibile dal MIT (Massachusetts Institute of Technology), che utilizza un KEMAR per determinare le HRIR a diverse angolazioni ed elevazioni.

È creato un codice MATLAB che svolga le funzioni di lettura, scrittura e elaborazione dei dati, ottenendo la possibilità di trasformare un qualunque file wave in un nuovo segnale audio le cui caratteristiche danno, all'ascolto, l'illusione che il suono provenga dalla prescelta direzione 3D.

# Bibliografia

[1] G. Bill, M. Keith, MIT Media Lab - *HRTF Measurements of a KEMAR Dummy-Head Microphone*, 1994:

<https://goo.gl/GyeOMu>

[2] Matthew Stapleton - *Matlab model for headphone based 3D sound from KEMAR based HRIR's*, University of Sidney:

<https://goo.gl/vUCorH>

[3] Xiao-li Zhong and Bo-sun Xie – *Head Related Transfer Functions and Virtual Auditory Display*, South China University of Technology, 2012:

<https://goo.gl/vKejRf>

[4] K.J. Faller, A. Barreto - *Simulation and Real-Time Implementation for Teaching 3D Sound*, Computers in Education Journal, Vol. XVI, 2006:

<https://goo.gl/bnMSRX>

[5] M. Stefanelli - *Audio Binaurale: Ascolto e Registrazione*:

<https://goo.gl/OqKb47>

[6] T. Potisk – *Head-Related Transfer Function*, Faculty of Mathematics and Physics, University of Ljubljana, 2015:

<https://goo.gl/SkwNfH>

[7] M. Geronazzo - *Modello strutturale di orecchio esterno per il rendering audio 3D*, Università di Padova, 2008

[8] E. Tavazzi - *Valutazione soggettiva di un modello di campo vicino per la percezione binaurale della distanza relativa tra sorgenti sonore virtuali*, Università di Padova, 2014

[9] G. Ricci, M. E. Valcher - *Segnali e Sistemi*, IV Edizione, 2010

# Appendice

## Script MATLAB

### - Hrtfmodel.m

```
function [ output ] = hrtfmodel(input, Azi, Ele, fs)
% La cartella HRIRData contiene misure effettuate tramite
% KEMAR per un insieme di angoli azimuth ed elevazione
% per filtrare un segnale di ingresso selezionato

% Variabili di ingresso:
% Input - File audio definito dall'utente
% Azi - Angolo di Azimuth
% Ele - Angolo di elevazione
% fs - Frequenza di campionamento

% Uscita:
% Output - HRTF
% NB: Un file wav stereo viene creato nella fase di uscita
% e compare nella cartella corrente con HRTFoutput
% seguita dalla elevazione e l'azimut.

%% Orecchio sinistro - Richiamo dei dati dal lato sinistro e
% elaborazione della sorgente di ingresso con la risposta
% all'impulso risultante
    filenameL =
    strcat('./HRIRData/elev',num2str(Ele),'/L',num2str(Ele),'e',
    num2str(Azi),'a.dat');
% Definire il nome del file da utilizzare basato sull' Azi
% e Ele

if Azi == 0
    filenameL =
    strcat('./HRIRData/elev',num2str(Ele),'/L',num2str(Ele),
    'e000a.dat'); % Se Azi è impostato su 0, tre 0 sono posti
% nel nome del file
```

```

elseif Azi > 0 && Azi < 10;
    filenameL =
strcat('./HRIRData/elev',num2str(Ele),'/L',num2str(Ele),'e00',
num2str(Azi),'a.dat'); % Se Azi è superiore a 0 e minore di
% 10, vengono aggiunti due 0
elseif Azi >= 10 && Azi < 99;
    filenameL =
strcat('./HRIRData/elev',num2str(Ele),'/L',num2str(Ele),'e0',
num2str(Azi),'a.dat'); % Se Azi è maggiore o uguale a 10 e
% fino a 99, si aggiunge un singolo 0
else
    filenameL =
strcat('./HRIRData/elev',num2str(Ele),'/L',num2str(Ele),'e',
num2str(Azi),'a.dat');
    % Per tutto il resto l'azimut può essere inserito
% direttamente nel nome del file
end

% Aprire il file definito e leggere i dati al suo interno
fpL = fopen(filenameL,'r','ieee-be'); % Apre il file
% come definito dalla funzione IF
dataL = fread(fpL, 256, 'short'); % Legge i dati forniti
% dal file
fclose(fpL); % Chiude nuovamente il file

leftimp = dataL(1:2:256); % La risposta all'impulso
% che era stata letta dal file

left = filter(leftimp,1,input); % La fase in cui il segnale
% di ingresso viene filtrato dalla risposta all'impulso
left = left./max(abs(left)); % Il segnale principale è
% diviso da solo per portare tutti i dati in un range di
% sicurezza

Left_Output = left; % La risposta di uscita a sinistra.

%% Orecchio destro - Richiamo dei dati dal lato sinistro e
% elaborazione della sorgente di ingresso con la risposta
% all'impulso risultante

```

```

        filenameR =
strcat('./HRIRData/elev',num2str(Ele),'/R',num2str(Ele),'e',
num2str(Azi),'a.dat');
% Definire il nome del file da utilizzare basato sull' Azi
% e Ele

if Azi == 0
    filenameR =
strcat('./HRIRData/elev',num2str(Ele),'/R',num2str(Ele),
'e000a.dat'); % Se Azi è impostato su 0, tre 0 sono posti
% nel nome del file

elseif Azi > 0 && Azi < 10;
    filenameR =
strcat('./HRIRData/elev',num2str(Ele),'/R',num2str(Ele),
'e00',num2str(Azi),'a.dat'); % Se Azi è superiore a 0 e
% minore di 10, vengono aggiunti due 0
elseif Azi >= 10 && Azi < 99;
    filenameR =
strcat('./HRIRData/elev',num2str(Ele),'/R',num2str(Ele),
'e0',num2str(Azi),'a.dat'); % Se Azi è maggiore o uguale
% a 10 e fino a 99, si aggiunge un singolo 0
else
    filenameR =
strcat('./HRIRData/elev',num2str(Ele),'/R',num2str(Ele),'e',
num2str(Azi),'a.dat');
% Per tutto il resto l'azimut può essere inserito
% direttamente nel nome del file
end

% Aprire il file definito e leggere i dati al suo interno
fpR = fopen(filenameR,'r','ieee-be'); % Apre il file
% come definito dalla funzione IF
dataR = fread(fpR, 256, 'short'); % Legge i dati forniti
% dal file
fclose(fpR); % Chiude nuovamente il file

rightimp = dataR(1:2:256); % La risposta all'impulso
% che era stata letta dal file

```



```

right = filter(rightimp, 1, input); % La fase in cui il
% segnale di ingresso viene filtrato dalla risposta all'impulso
right = right./max(abs(right)); % Il segnale principale
% è diviso da solo per portare tutti i dati in un range di
% sicurezza

Right_Output = right; % La risposta di uscita a sinistra.
%% Somma di uscita e il file wav di produzione
output = [Left_Output Right_Output]; % Somma stereo dei
% due file di uscita

wavwrite(output, fs, 32,
strcat('HRTFOutput',num2str(Ele),'e',num2str(Azi),'a'));
% Creazione di un file wav del segnale di uscita risultante
end

```

- Sample Stereo Movement.m

```

input = audioread('corvaglia.wav');

fs = 44100; % frequenza di campionamento del file audio
% definito dall'utente
ele = 10;

s1=hrtfmodel(input, 45, ele, fs);

s2=hrtfmodel(input, 120, ele, fs);

s3=hrtfmodel(input, 270, ele, fs);

s4=hrtfmodel(input, 320, ele, fs);

s=vertcat(s1,s2,s3,s4); % Concatena le matrici
% verticalmente per unire l'audio a diverse angolazioni

sound(s,fs); % Invia il segnale audio s
% all'altoparlante alla frequenza di campionamento fs

disp(s); % Mostra matrice delle s

```

## **Ringraziamenti**

Questa tesi di laurea è il raggiungimento di uno dei più importanti obiettivi della mia vita.

Essere qui a scrivere i ringraziamenti è una grande vittoria!

Desidero ringraziare quindi tutti coloro che hanno permesso e favorito il raggiungimento di tale obiettivo. Spero di non dimenticare nessuno, ma se tu mio caro lettore, non dovessi trovare il tuo nome nelle prossime righe, dimmelo e ti ringrazierò di persona, anche per il solo fatto che tu abbia interesse a leggere.

In questo mio cammino desidero ringraziare il Professore Coluccia, relatore di questa tesi, per la grandissima disponibilità e cortesia dimostratemi e per tutto l'aiuto fornito durante la stesura.

Un ringraziamento ai miei genitori che con il loro sostegno morale ed economico mi hanno permesso di raggiungere questo traguardo importante; alle mie nonne, a mio fratello Giacomo, ai miei zii, cugini, alla mia ragazza Fiorella e a tutti i miei amici più cari.

E infine un grazie ai miei compagni di studi, in particolar modo Emanuele, Matteo, Franco e Gabriele per avere condiviso momenti facili e difficili; loro sono stati per me non solo semplici compagni di studio, ma soprattutto veri amici.

**GRAZIE nonno Salvatore e nonno Giacomo**