

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

CAMPUS DI CESENA

SCUOLA DI INGEGNERIA E ARCHITETTURA

CORSO DI LAUREA IN INGEGNERIA BIOMEDICA

Tesi di Laurea:

Elaborazione di segnali audio per la localizzazione
di sorgenti

Elaborato in:

Elaborazione dei segnali

Relatore:

Chiar.mo Prof. Marco Chiani

Correlatore:

Dott. Enrico Paolini

Presentata da:

Camilla Urbani

Sessione II

Anno accademico 2014/2015

Abstract

In questo lavoro viene trattata l'elaborazione di segnali audio per la localizzazione di sorgenti sonore. Di certo gli elementi che concorrono maggiormente nella localizzazione del suono sono le nostre orecchie: esse sono separate dalla testa, che funge da ostacolo acustico.

E' importante distinguere se la sorgente si trova davanti agli occhi dell'ascoltatore, ossia nel piano interaurale, o se altrimenti occupa una posizione più o meno laterale.

In tali due casi i segnali audio raggiungeranno in modo diverso le due orecchie: avremo rispettivamente che essi, in un caso, raggiungeranno le due orecchie contemporaneamente; nell'altro una delle due orecchie riceverà un segnale ritardato ed attenuato.

Tale elaborato si propone di ricreare uno scenario di classe virtuale in cui, attraverso l'utilizzo di MATLAB, vengono riprodotti i vari segnali cercando di fornire la sensazione di un ascolto reale, andando ad operare su di essi.

Indice

Introduzione	1
1 Base fisica e psicofisica della localizzazione sonora	2
1.1 L'orecchio e la percezione del suono	2
1.1.1 Il padiglione auricolare come codificatore direzionale del suono	3
1.1.2 Il canale uditivo	4
1.1.3 L'orecchio medio	4
1.1.4 L'orecchio interno	6
1.2 Localizzazione del suono: l'audio binaurale	8
1.3 La teoria Duplex di Lord Rayleigh	11
1.4 HRIR e HRTF	15
1.4.1 Contributo alla HRTF delle diverse componenti della periferia uditiva	17
1.4.2 Tecniche di misurazione	18
1.4.3 Misurazioni delle HRTF usando tubi sonda	19
1.4.4 Sistema di coordinate	19
1.5 La testa come ostacolo acustico	21
1.6 Segnali che si manifestano a causa della differenza nella lunghezza di percorso	24
1.7 Individuazione della distanza della sorgente	27
1.8 Individuazione dell'elevazione della sorgente	28

1.9 Sensibilità psicofisica alle onde acustiche per una localizzazione del suono	29
1.9.1 Sensibilità alle differenze di tempo interaurale	29
1.9.1.1 Effetto Haas o "effetto precedenza"	30
1.9.1.2 Misure neurofisiologiche della sensibilità relativa alla ITD	32
1.9.2 Sensibilità alle differenze di livello interaurale	34
2 Virtual Classroom	35
2.1 Analisi dello scenario di un tipico ambiente di classe	35
2.2 Riproduzione con matlab di un ambiente di ascolto virtuale	38
Riferimenti	45

Introduzione

La tesi ha l'obiettivo principale di analizzare l'aspetto fisico e psicofisico del processo di localizzazione di una sorgente sonora, successivamente sfruttate in un processo di riproduzione di un tipico ambiente di classe virtuale, ricreato attraverso MATLAB, con l'obiettivo di fornire una reale sensazione sonora.

Lo studio della sintesi dell'audio binaurale si protrae ormai da un centinaio di anni: nel XX secolo, il fisico Lord Rayleigh derivò una formula matematica per la diffrazione, intorno ad una sfera rigida, di un'onda sonora prodotta da una sorgente a distanza infinita.

Successivamente, si è analizzato come l'uomo possa riuscire a discriminare la direzione di una sorgente ed in seguito la percezione che egli ha della distanza della stessa.

Nel primo capitolo, verranno trattate le componenti dell'orecchio seguite dai vari fattori e dalle varie teorie ed osservazioni che sono alla base della localizzazione della sorgente sonora.

Nel secondo capitolo viene presentata una prima analisi di uno scenario di classe virtuale e successivamente, attraverso MATLAB, viene fornito un primo modello per la localizzazione della sorgente, tralasciando l'effetto filtro dell'orecchio esterno, l'effetto schermante della testa ed i contributi relativi all'elevazione della sorgente sonora.

Capitolo 1

Base fisica e psicofisica della localizzazione sonora

1.1 L'orecchio e la percezione del suono

L'organo fondamentale per la percezione del suono è l'orecchio: esso agisce da trasduttore nel trasformare l'energia acustica prima in energia meccanica e, successivamente, in energia elettrica inviata al cervello per l'elaborazione tramite le terminazioni nervose.

Le tre parti dell'apparato uditivo umano sono: l'orecchio esterno, l'orecchio medio e l'orecchio interno [1].

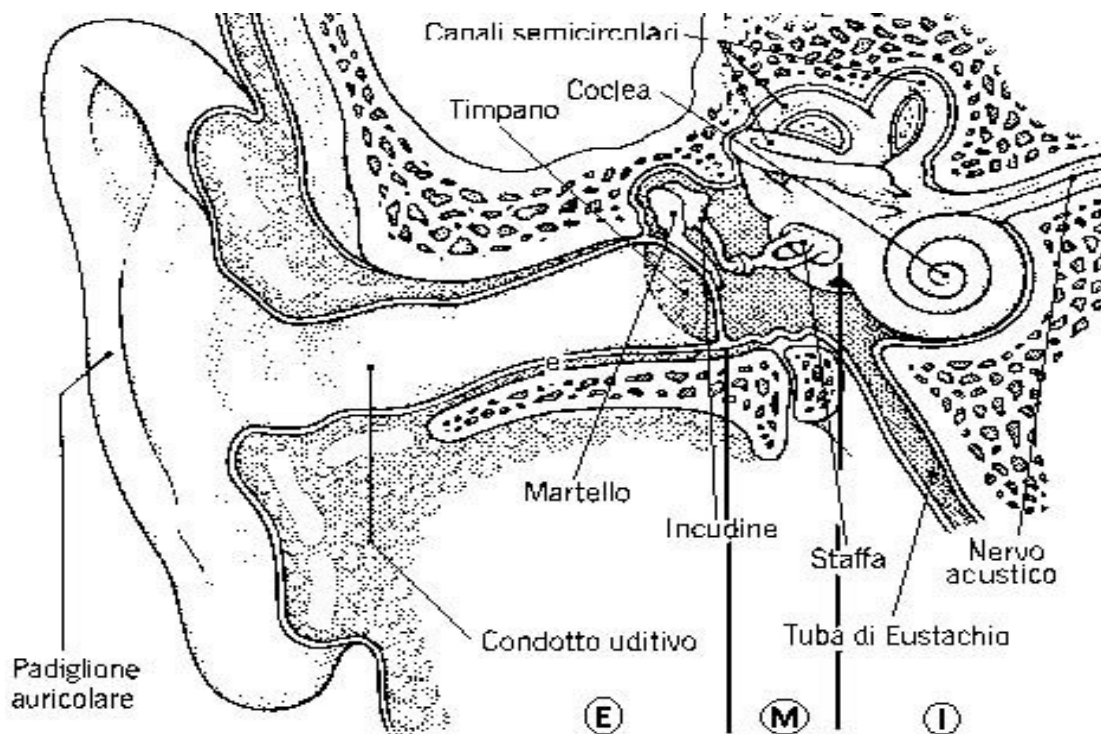


Figura 1.1: Le quattro parti principali dell'orecchio umano: il padiglione auricolare, il canale uditivo, l'orecchio medio e l'orecchio interno.

L'orecchio esterno è costituito dal padiglione auricolare e dal canale uditivo. Il canale uditivo termina con la membrana del timpano.

L'orecchio medio è formato da una cavità riempita d'aria in cui si trovano tre ossicini chiamati rispettivamente martello, incudine e staffa. Il martello è unito al timpano e la staffa è attaccata alla finestra ovale dell'orecchio interno. Questi tre ossicini nel loro insieme formano un collegamento meccanico a leva fra il timpano, attivato dalla percussione delle particelle d'aria, e la chiocciola (o coclea), situata nell'orecchio interno.

Quest'ultimo termina con il nervo acustico, che invia impulsi al cervello.

L'onda sonora che investe l'orecchio dà luogo a movimenti meccanici che generano scariche nervose che raggiungono il cervello, creando una sensazione.

La domanda a cui ancora oggi risulta difficile dare una risposta è: " Come vengono riconosciuti ed interpretati questi suoni? "

1.1.1 Il padiglione auricolare come codificatore direzionale del suono

In passato, il padiglione auricolare veniva considerato come un semplice dispositivo il cui ruolo era quello di raccogliere il suono [1].

Ricerche recenti hanno rivelato che il padiglione svolge un ruolo fondamentale nella valutazione della direzionalità di tutti i suoni percepiti dall'orecchio.

La pinna, ovvero il padiglione auricolare, offre una vasta superficie al fronte sonoro e permette di raccoglierne un'ampia porzione; il suono viene riflesso dal padiglione auricolare e concentrato verso il condotto uditivo.

Essa è fondamentale per la localizzazione del suono nello spazio, apportando importanti informazioni circa l'elevazione della sorgente sonora.

Le informazioni relative alla direzione in cui si trova la sorgente vengono sovrainpresse al contenuto del suono stesso, in modo che la pressione sonora risultante al timpano permetta al cervello di interpretare sia il contenuto del messaggio sonoro, sia la direzione di provenienza.

1.1.2 Il canale uditivo

Anche il canale uditivo contribuisce ad aumentare la sensazione dei suoni che lo attraversano [1].

Il canale uditivo può essere semplificato con un cilindro di diametro di circa 0.7 cm e lunghezza di circa 3 cm: è un condotto a forma di tubo che all'estremità interna termina con la membrana del timpano.

La risonanza ad alcune frequenze aumenta la pressione sonora sul timpano, l'effetto massimo si ha per frequenze attorno ai 3000 Hz: ad esse corrispondono delle lunghezze d'onda di circa 12 cm e quindi i 3 cm del canale rappresentano 1/4 della lunghezza d'onda.

1.1.3 L'orecchio medio

La trasmissione dell'energia sonora da un mezzo rarefatto come l'aria ad un mezzo denso come l'acqua, non è semplice.

Nel caso dell'udito umano, si tratta di riuscire a trasferire con la massima efficienza al fluido dell'orecchio interno la fievole energia associata al moto vibratorio di un sottile diaframma [1].

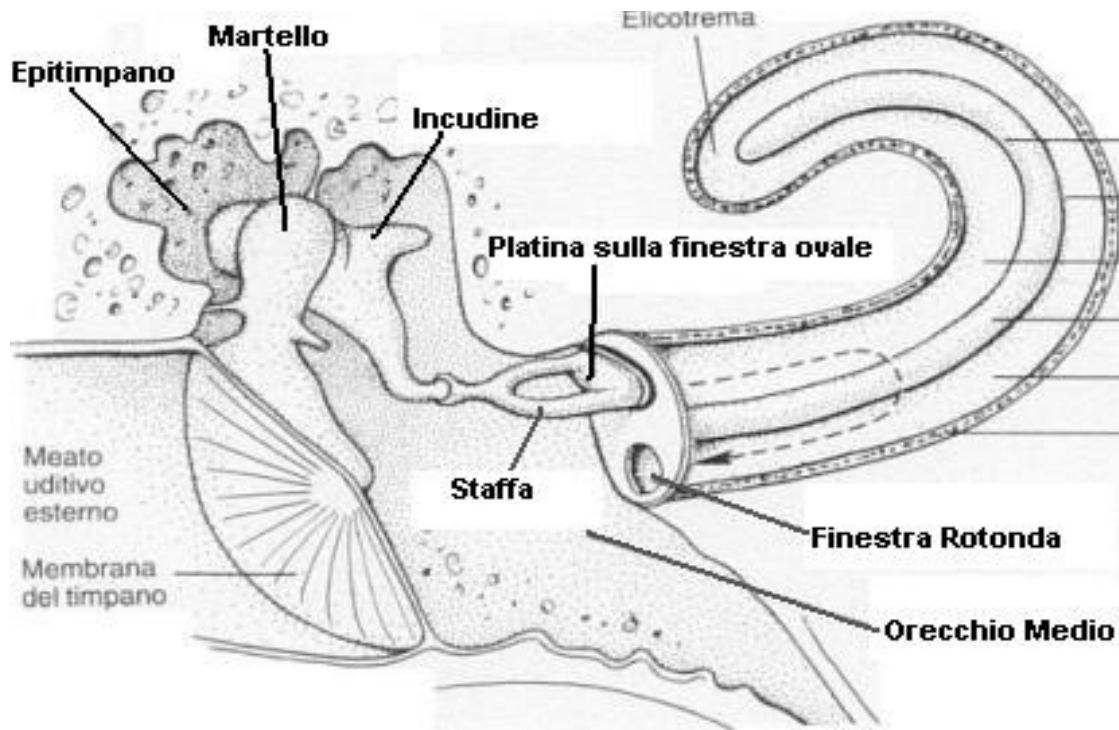


Figura 1.1.3: orecchio medio.

Si tratta di un problema di adattamento di impedenze acustiche.

I tre ossicini (martello, incudine e staffa), formano un sistema di collegamento meccanico tra il timpano e la finestra ovale, che è in diretto contatto con il fluido dell'orecchio interno.

Il martello è attaccato al timpano mentre la staffa può essere considerata una parte della finestra ovale.

In questo collegamento, lo spostamento del timpano è ridotto quando esso raggiunge la finestra ovale. Questo è solo un contributo all'adattamento di impedenza acustica; l'altro deriva dalle aree di certe superfici. L'area del timpano è di circa 80 mm^2 mentre l'area della finestra ovale è di circa 3 mm^2 ; pertanto una determinata forza applicata al timpano viene ridotta secondo il rapporto $80/3$, ovvero di circa 27 volte.

L'adattamento delle impedenze insieme all'amplificazione introdotta dalla risonanza, sono efficaci: un movimento del diaframma dell'ordine delle dimensioni molecolari è in grado di generare una percezione, anche se al limite della soglia di udibilità.

La finestra rotonda separa l'orecchio medio, riempito d'aria, dal fluido praticamente incompressibile dell'orecchio interno.

La tromba di Eustachio, che segue la finestra rotonda, consente di equilibrare la pressione statica dell'aria dell'orecchio medio con la pressione atmosferica esterna, in modo che il timpano e le membrane dell'orecchio interno possano funzionare a dovere. Ogni volta che deglutiamo, la tromba di Eustachio si apre, equilibrando la pressione dell'orecchio medio.

1.1.4 L' orecchio interno

Risulta molto più problematico comprendere le complesse funzioni della coclea o chiocciola, ancora non del tutto spiegate [1].

Le vibrazioni del timpano attivano gli ossicini. Il movimento della staffa, che è solidale con la finestra ovale, fa vibrare il fluido dell'orecchio interno. Uno spostamento verso l'interno della finestra ovale, determina un movimento del fluido che causa lo spostamento verso l'esterno della finestra rotonda.

Il suono che attiva la finestra ovale dà luogo alla formazione di onde stazionarie sulla membrana basilare.

La membrana basilare ospita una popolazione di cellule acustiche cigliate, circa 4000, che vibrano in accordo con la vibrazione del fluido. Ogni gruppo di ciglia è collegato ad una terminazione nervosa in grado di convertire la vibrazione ricevuta dal fluido in impulsi elettrici percepiti dal cervello come suoni. La posizione del picco di ampiezza dell'onda stazionaria sulla membrana basilare cambia secondo la frequenza del suono eccitante.

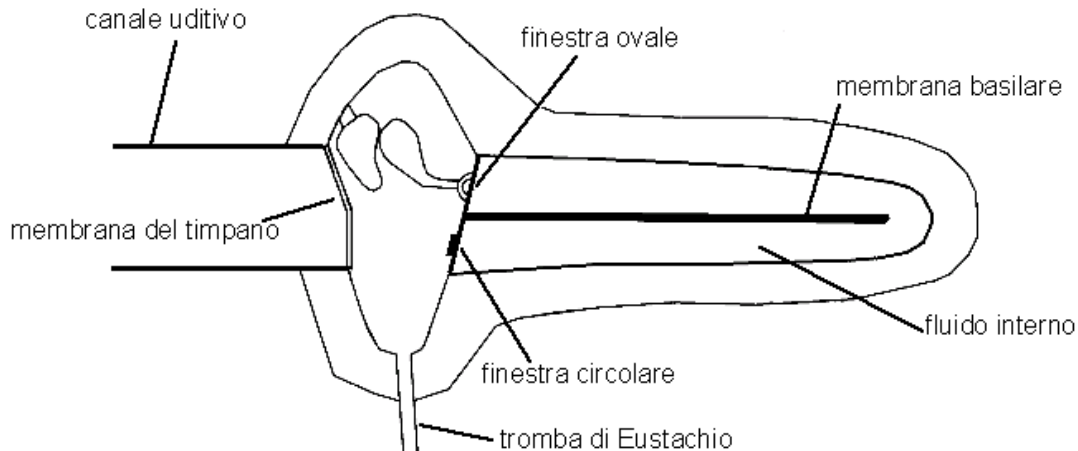


Figura 1.1.4: orecchio interno.

Se il suono è a bassa frequenza, il picco di ampiezza massima si manifesta in prossimità dell'estremità lontana della membrana basilare; se invece è ad alta frequenza, si manifesta in vicinanza della finestra ovale.

Per segnali complessi, quali la voce umana o la musica, vengono prodotti molti picchi temporanei, che variano in continuazione per ampiezza e posizione lungo la membrana basilare. Tali picchi di risonanza sono molto stretti per basse intensità di suono, mentre si allargano per i suoni più intensi.

Le onde formate sulla membrana basilare, stimolano i numerosissimi terminali nervosi ciliati che trasmettono segnali al cervello in forma di scariche nervose.

Quando un suono eccita il fluido dell'orecchio interno, la membrana e le cellule ciliate sono stimulate e inviano un'onda elettrica attraverso il tessuto circostante; tale onda è descritta in termini di potenziali microfonici (analogici), che possono essere raccolti e amplificati, riproducendo così l'impatto del suono sull'orecchio, che funziona quindi come un microfono biologico. Tali potenziali sono proporzionali alla pressione sonora e danno luogo a una risposta lineare su un range di 80 dB. I potenziali microfonici non devono essere confusi con i potenziali di azione del nervo uditivo, che trasmettono le informazioni al cervello.

Mentre i segnali microfonici sono analogici, gli impulsi inviati alla corteccia uditiva sono generati da scariche di neuroni. Una singola fibra

nervosa è attiva o disattiva e quando una viene attivata, essa va ad attivare anche quella adiacente e così via.

Presumibilmente, il volume del suono dipende invece dal numero di fibre nervose eccitate e dal tasso di ripetizione di tali eccitazioni. La soglia di sensibilità corrisponde all'eccitazione di una fibra.

1.2 Localizzazione del suono: l'audio binaurale

L'ambiente nel quale un suono viene riprodotto modifica ciò che percepiamo attraverso un fenomeno detto riverberazione. Esso è legato alla riflessione del suono da parte di ostacoli posti nell'ambiente in cui viene riprodotto e causa la propagazione di versioni ritardate o attenuate del suono stesso.

Gli elementi che modificano il suono emesso dalla sorgente da quello percepito sono, tra gli altri:

- il busto e le spalle;
- la testa del soggetto;
- la pinna dell'orecchio, ovvero il padiglione auricolare.

Questi elementi sono, con differenze più o meno evidenti, diversi da una persona all'altra quindi un suono proveniente da una singola sorgente può essere udito in versioni diverse da soggetti distinti. Busto, testa e ambiente contribuiscono nella modifica del suono e una particolare tecnica di riproduzione audio, l'audio virtualizzato, ha lo scopo di presentare ai due timpani delle orecchie gli stessi segnali che produrrebbe un suono reale posizionato in quel punto nello spazio.

La percezione della direzione da cui proviene il suono è, per lo meno in parte, il risultato della sorprendente funzione di codifica del padiglione auricolare.

Il suono riflesso dalle varie cavità, nervature e superfici del padiglione

auricolare si combina con il suono non riflesso (suono diretto), all'entrata del canale uditivo.

Questa combinazione, codificata con le informazioni direzionali, passa, come precedentemente descritto, per il canale uditivo fino al timpano e da lì all' orecchio medio e a quello interno, per poi giungere al cervello per l'interpretazione.

In passato, si riteneva che il fatto di avere due orecchie fosse analogo alla presenza di due polmoni o di due reni: se uno dei due organi gemelli avesse subito qualche danno, l'altro avrebbe potuto continuare a funzionare.

Successivamente, questa visione venne superata e si mise in evidenza il fatto che le due orecchie collaborano per la localizzazione binaurale.

I fattori coinvolti sono: la differenza di livello e la differenza di tempo di arrivo (fase) dei suoni che investono le due orecchie.

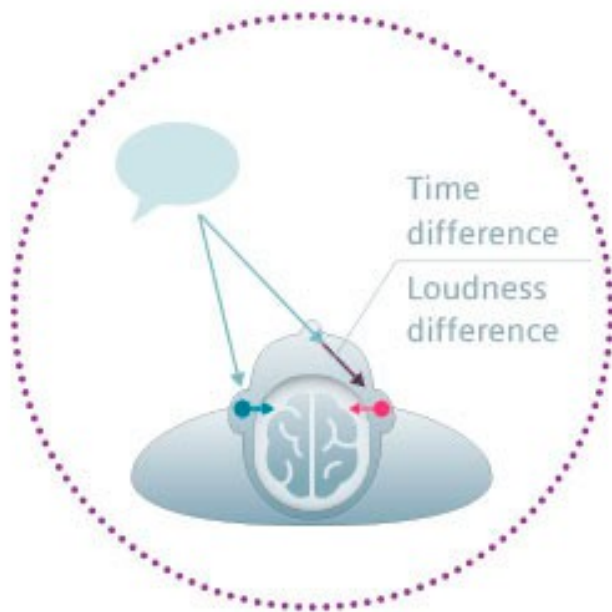


Figura 1.2.1: il nostro senso direzionale binaurale dipende dalle differenze di livello e di fase del suono che investe le due orecchie.

L'orecchio più vicino alla sorgente riceve un segnale di intensità maggiore rispetto all'altro, in quanto il cranio forma una sorta di "ombra sonora". A causa della differenza di distanza dalla sorgente, l'orecchio più lontano riceve il suono con un certo ritardo rispetto a quello più vicino.

Esiste anche un punto cieco di localizzazione: l'ascoltatore non è in grado di dire se i suoni provengano direttamente dal fronte o dal retro, poiché l'intensità del suono che giunge ad ogni orecchio è la stessa e possiede la stessa fase.

E' inoltre importante sottolineare che il suono che raggiunge in modo "diretto" l'ascoltatore, crea in quest'ultimo la principale percezione della direzione e si parla quindi di *legge del primo fronte d'onda*.

Nella localizzazione sonora, dobbiamo inoltre evidenziare i seguenti aspetti fondamentali:

1. il ruolo importante delle informazioni disponibili ad ogni orecchio (segnali monaurali per la localizzazione del suono);
2. l'analisi nel dominio della frequenza (segnali spettrali monaurali e binaurali);
3. le caratteristiche del suono nel dominio del tempo.



A)



B)

Figura 1.2.2: A) registratore digitale con i rispettivi microfoni omnidirezionali; B) testa artificiale della Sennheiser generalmente sfruttata nei teatri (posta all' altezza di 1.20 m, in modo da simulare la presenza di uno spettatore seduto). I microfoni, posizionati nelle orecchie artificiali sono omnidirezionali in modo da ricreare una registrazione binaurale; sono diretti verso il timpano in modo da captare al meglio come il segnale arriva all'orecchio umano. Ad essi è collegato il preamplificatore.

1.3 La teoria Duplex di Lord Rayleigh

La principale affermazione teoretica alla base della localizzazione del suono nel piano azimutale è conosciuta come "Duplex Theory" che ha le sue basi nel lavoro di Lord Rayleigh nel 1907.

Essa si basa sul fatto che la principale differenza tra le due orecchie è che essi non si trovano nella stessa posizione.

I primi studi erano basati su osservazioni fisiche e psicofisiche piuttosto elementari. I modelli del comportamento delle onde sonore attorno alla

testa venivano eseguiti considerando la testa come una sfera e le due orecchie come due punti di ricevimento collocati simmetricamente. Per la maggior parte del secolo, i modelli risultanti da questa semplificazione, hanno dominato il programma della ricerca.

Il fatto che noi abbiamo due orecchie separate da una testa relativamente larga significa che, per i suoni fuori dalla linea mediana, vi sono delle differenze nelle lunghezze del percorso dalla sorgente sonora a ciascun orecchio. Questo comporta una differenza nel tempo di arrivo del suono a ciascuno di essi e nel livello sonoro con cui esso si manifesta a ciascun orecchio dopo aver attraversato l'ostacolo rappresentato dalla testa.

Si fa riferimento a queste grandezze definendole **"interaural time difference (ITD)"** e **"interaural level difference (ILD)"**.

L' ITD si manifesta nella differenza di inizio del suono ad ogni orecchio e, per suoni continui, si traduce in una differenza interaurale nella fase dei suoni ad ogni orecchio **"interaural phase difference (IPD)"**.

Psicofisicamente, a frequenze al di sotto di circa 1,5 kHz, si ha che la lunghezza d'onda di un ipotetico segnale sinusoidale, risulta maggiore del diametro della testa; lo scostamento laterale percepito, sarà proporzionale alla differenza di fase del suono presente alle due orecchie (interaural phase difference).

Poichè la testa è un mezzo relativamente denso, tenderà a riflettere e a rifrangere le onde sonore. Questo costituisce un effetto significativo quando le lunghezze dell'onda sonora sono dello stesso ordine o più piccole del diametro della testa (distanza tra i due orecchi: circa 18 cm) .

Per un suono localizzato fuori la linea mediana, la testa getta un'ombra acustica per l'orecchio lontano e genera una differenza interaurale nel livello del suono per ogni orecchio (interaural level difference (ILD)).

A basse frequenze di ascolto questo effetto è trascurabile perchè la lunghezza delle onde coinvolte è relativamente lunga, ma per le

frequenze al di sopra di circa 1.5 kHz, l'importanza dell'effetto risulta notevole.

L'ammontare dell'ombreggiatura dipenderà dalla posizione della sorgente: meno energia raggiunge l'orecchio lontano, più distante è situata la fonte sonora. Ci sono anche dei cambiamenti nel livello del suono per l'orecchio più vicino alla sorgente, anch'essi dipendenti dalla posizione della stessa.

Queste ultime variazioni nel livello, risultano da due distinti effetti:

-il cosiddetto effetto ostacolo (paragrafo 1.5);

-effetto filtro dell'orecchio esterno (paragrafo 1.4).

Si può quindi affermare che in generale per frequenze al di sotto di circa 1.5 kHz, prevale l'effetto di fase (tempo), mentre per frequenze al di sopra di circa 1.5 kHz, prevale l'effetto del livello.

La "Duplex Theory" è incompleta: non spiega in alcun modo come possa essere determinata la posizione del suono al di fuori del piano azimutale; inoltre, a causa della sistemazione geometrica delle orecchie, una singola differenza interaurale nel tempo o nel livello non viene associata con una singola localizzazione spaziale: una particolare differenza interaurale, definirà la superficie di un cono immaginario centrato sull'asse interaurale, il cosiddetto cono di confusione [2].

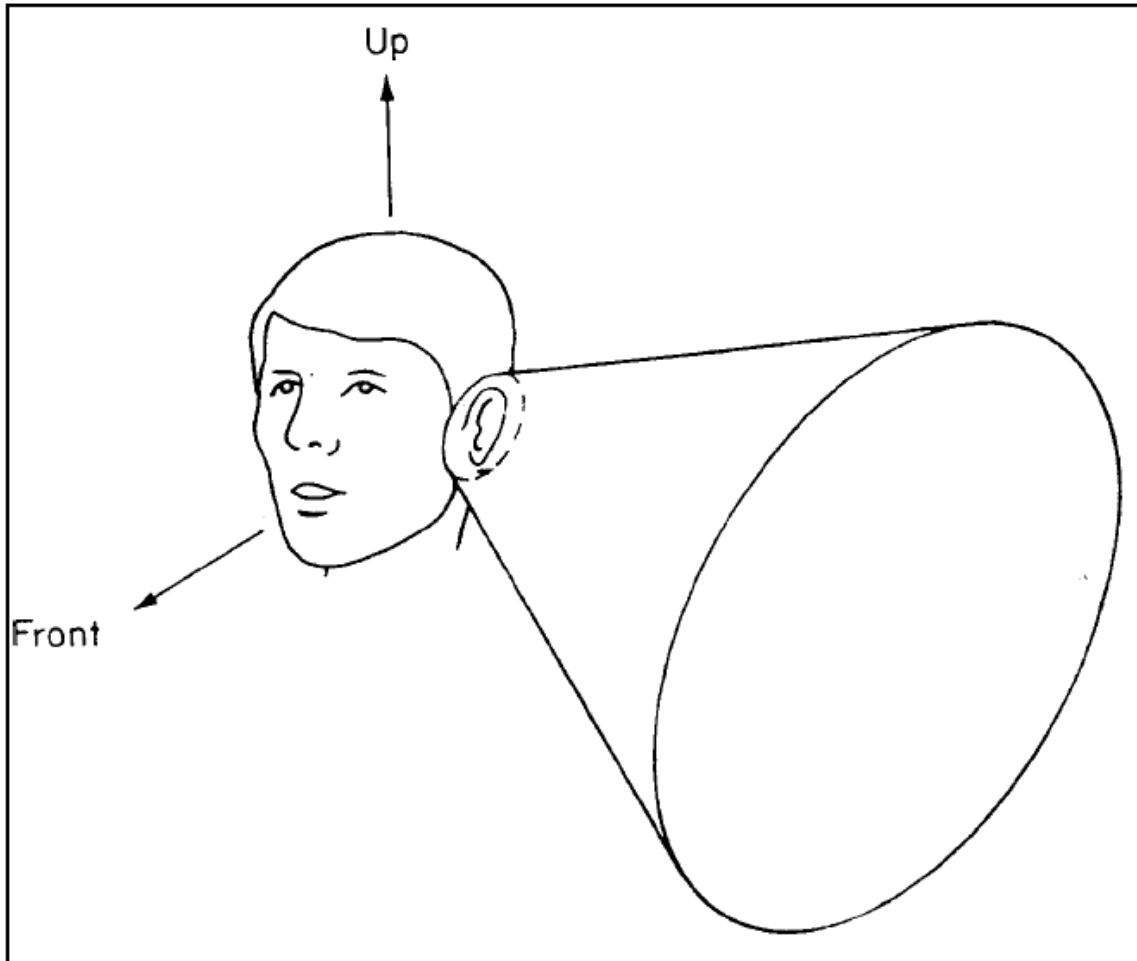


Figura 1.3: cono di confusione [2].

L'angolo solido del cono verrà associato con l'ampiezza dell'intervallo; per esempio il cono diviene il piano mediano per una differenza di tempo interaurale uguale a zero, e diviene l'asse interaurale per una differenza di tempo interaurale massima, a prescindere dalla distanza della sorgente.

Quando si cerca di individuare una sorgente sonora che si trova di fronte (0°) o alle nostre spalle (180°), poichè le differenze interaurali risultano minime sul piano mediano, ci troviamo in uno stato confusionale. Questo fenomeno viene appunto descritto come: "**cono di confusione**".

L'incompletezza della teoria duplex è illustrata anche dal fatto che ascoltatori sordi da un orecchio, possono localizzare il suono con un

notevole grado di accuratezza. Ciò suggerisce che segnali fisici, piuttosto che il tempo interaurale e le differenze di livello, vengono usati nella localizzazione del suono. Tali ulteriori informazioni vanno ricercate quindi nello spettro del segnale. Purtroppo la relazione tra lo spettro del segnale e la locazione spaziale del suono non è semplice. Per analizzare questo problema, molti ricercatori hanno effettuato delle misurazioni dei segnali presenti agli orecchi degli ascoltatori per diverse posizioni spaziali per un dato suono. Queste misurazioni sono dette HRTF e riassumono il filtraggio acustico operato dalla testa, dal torso e dal padiglione auricolare, dipendente dalla posizione spaziale della sorgente sonora.

1.4 HRIR e HRTF

La pressione sonora prodotta da una sorgente nel canale uditivo, è determinata esclusivamente dalla risposta impulsiva nel tragitto dell'onda sonora dalla sorgente stessa al timpano, chiamata HRIR (head-related-impulse-response).

Formalmente, una singola HRIR è definita come una risposta impulsiva di un singolo orecchio (destro o sinistro) a un suono, misurata da uno specifico punto nello spazio. La convoluzione del segnale emesso da una sorgente arbitraria ($x(t)$) con la HRIR specifica della posizione, converte il suono emesso dalla sorgente in quello percepito dall'utente con il suo sistema uditivo ($y(t)$):

$$y(t) = x(t) * \text{HRIR}$$

La sua trasformata di Fourier, è detta HRTF (head-related-transfer-function) e contiene tutte le informazioni istantanee riguardanti la trasformazione dell'onda sonora nello spazio circostante. Esse raccolgono le informazioni su come l'orecchio umano "filtra" il suono proveniente dalle varie direzioni.

Chiaramente, esiste una HRTF per ogni orecchio. Queste due HRTF sono diverse tra loro.

Essendo la HRTF la trasformata di una risposta impulsiva, essa è funzione di trasferimento e pertanto tra output e input sussiste la relazione:

$$\text{HRTF} = Y(f) / X(f)$$

Ad influenzare l'HRTF, oltre al mezzo in cui le onde si propagano, sono gli effetti procurati dalle varie zone del corpo dell'ascoltatore: in particolare la testa, le orecchie, il torso e le spalle.

La testa fa sì che l'onda sonora sia diffratta tutt'intorno ad essa ed ha un effetto schermante sulle onde ad alta frequenza.

L'orecchio esterno, invece, possiede delle caratteristiche zone di risonanza che attenuano o amplificano una o più particolari frequenze; inoltre, riflette le onde sonore ad alta frequenza (dunque con una lunghezza d'onda relativamente limitata) a seconda della loro direzione.

Infine, le spalle ed il torso generano un'ombra sui suoni provenienti da sotto l'ascoltatore ed aggiungono ulteriori riflessioni che si sovrappongono al suono diretto.

Tutti gli effetti causati dalle varie parti del corpo sono lineari e si combinano in maniera addittiva nella HRTF complessiva.

Dunque, i comportamenti della testa, dell'orecchio esterno e di spalle e torso possono essere analizzati separatamente, sintetizzati e quindi sommati nella funzione di trasferimento complessiva.

Questo è l'approccio seguito nel progetto dei *modelli strutturali*.

1.4.1 Contributo alla HRTF delle diverse componenti della periferia uditiva

Considerando le funzioni di trasferimento spettrali registrate ad ogni estremità del canale uditivo, è importante ricordare che alcune strutture oltre al padiglione auricolare, contribuiranno a queste funzioni. La figura 1.4.1 mostra il relativo contributo di varie componenti della periferia uditiva calcolato per un suono collocato a 45° azimuth [2].

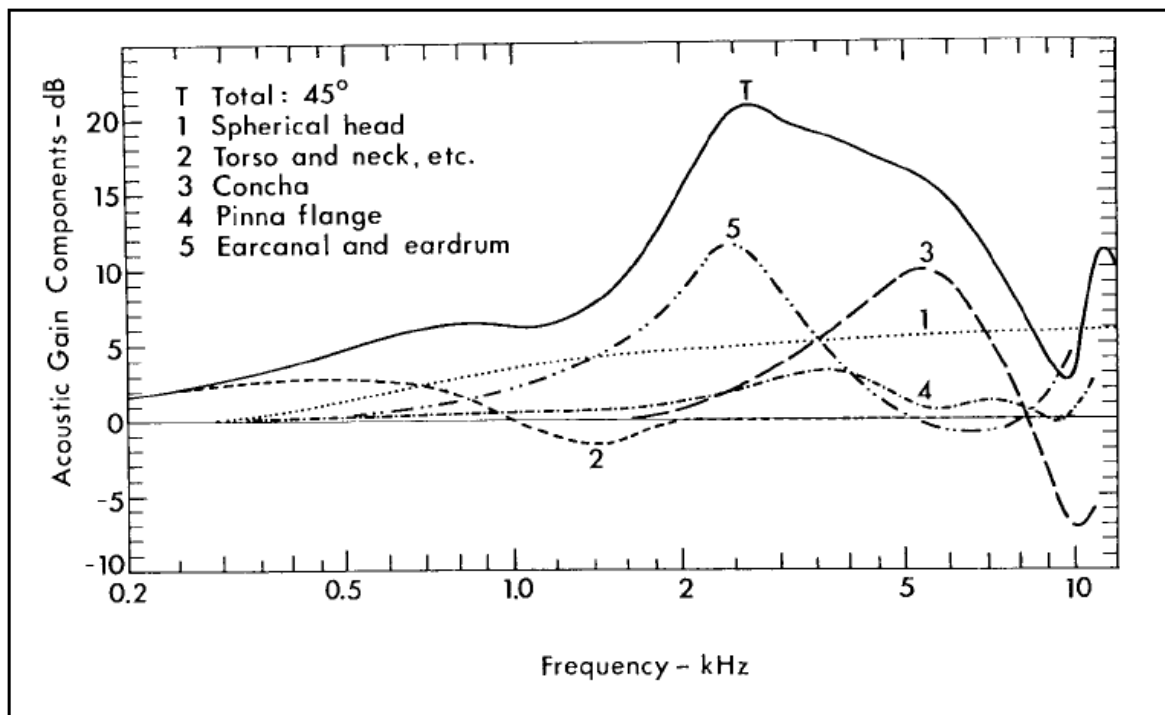


Figura 1.4.1: contributi delle diverse componenti della periferia uditiva, calcolati da Shaw (1974), rispetto ad una sorgente situata a 45° dal piano mediano [2].

Queste misure sono di prima approssimazione, ma servono per illustrare il punto che le caratteristiche della HRTF dipendono da un numero di diverse strutture fisiche.

Il guadagno dovuto alla testa, calcolato dalla descrizione di Rayleigh-Stewart della distribuzione della pressione sonora attorno ad una sfera, aumenta con frequenza crescente fino ad un asintoto di 6 dB. L'indice di

tale aumento, come una funzione di frequenza, è determinato dal raggio della sfera.

Negli umani, questo corrisponde ad un raggio di 8.75 cm e il punto mediano all'asintoto si presenta a 630 Hz (figura 1.5).

Il contributo del dorso e del collo è piccolo e circoscritto soprattutto alle basse frequenze.

Questi cambi di pressione, probabilmente risultano dalle interazioni delle onde sonore distribuite all'altezza dell'orecchio e sono efficaci soprattutto per le basse frequenze.

Il contributo del padiglione auricolare è piccolo a 45° azimuth, ma probabilmente esercita un'influenza più grande sul risultato totale per i suoni presentati dietro l'asse interaurale.

Il contributo più grande è attribuibile alla conca ed al complesso: canale uditivo-timpano.

Un importante aspetto di questi contributi è la complementarità dei componenti della conca e del canale uditivo che agiscono insieme per produrre un guadagno sostanziale su un'ampia gamma di frequenze.

Comunque un'importante distinzione tra i due è che il contributo del canale uditivo non è sensibile alla localizzazione dello stimolo, mentre il guadagno dovuto alla conca e alla sporgenza della pinna è chiaramente dipendente dalla direzione dello stimolo.

Quindi, l'HRTF è chiaramente composta sia da componenti che dipendono dalla collocazione che da componenti indipendenti dalla collocazione.

1.4.2 Tecniche di misurazione

Per il calcolo della HRTF, vi sono due metodi principali:

- la testa può essere trattata come una sfera rigida di raggio variabile a seconda del soggetto: tale assunzione è l'unica che consente di procedere per via analitica, ma è anche la meno precisa;

- attraverso l'utilizzo di microfoni, posizionati su un manichino: le misure sono effettuate in camere anecoiche e per un numero elevato di locazioni spaziali.

Le HRTF sono misurate a diversi azimuth (direzione destra – sinistra), a diversi gradi di elevazione (direzioni alto-basso) e ad una certa distanza dalla testa del manichino, in generale [2].

Il risultato finale di tali misure, consiste in un set di funzioni di trasferimento direzionale, comunemente detto HRTF.

1.4.3 Misurazioni delle HRTF usando tubi sonda

Per stabilire le trasformazioni spettrali dei suoni sono stati usati i toni puri e l'analisi di Fourier delle risposte all'impulso sia nei piani orizzontali che in quelli verticali.

Recentemente, sono state registrate le HRTF per ciascun orecchio per circa 350 localizzazioni nello spazio e si sono esaminati i cambiamenti nelle HRTF per le variazioni in azimuth e in elevazione sulla linea mediana anteriore e sull'asse interaurale [2].

Queste registrazioni sono state ottenute usando tubi sonda collocati a 6 mm dal timpano usando una tecnica acustica per assicurare accuratezza nella collocazione.

Usando la testa di un manichino dotata di microfono interno si sono anche calibrate le perturbazioni acustiche del sistema di registrazione.

1.4.4 Sistema di coordinate

Specificare la precisa collocazione di uno stimolo sonoro richiede l'adozione di un sistema di coordinate che per lo meno descriva lo spazio bidimensionale attorno al soggetto [2].

La forma più comune di sistema di coordinate è **il sistema polo singolo**, lo stesso sistema che viene usato per specificare la collocazione sulla superficie del pianeta.

Con la testa al centro di una sfera immaginaria, la collocazione azimuth è specificata dalle linee della latitudine dove direttamente in avanti di solito si stabilisce che si trovi l'azimuth 0° , con collocazioni negative alla destra della linea mediana anteriore.

L'elevazione è specificata da linee di longitudine con l'orizzonte audiovisivo a 0° di riferimento e le elevazioni inferiori che sono negative.

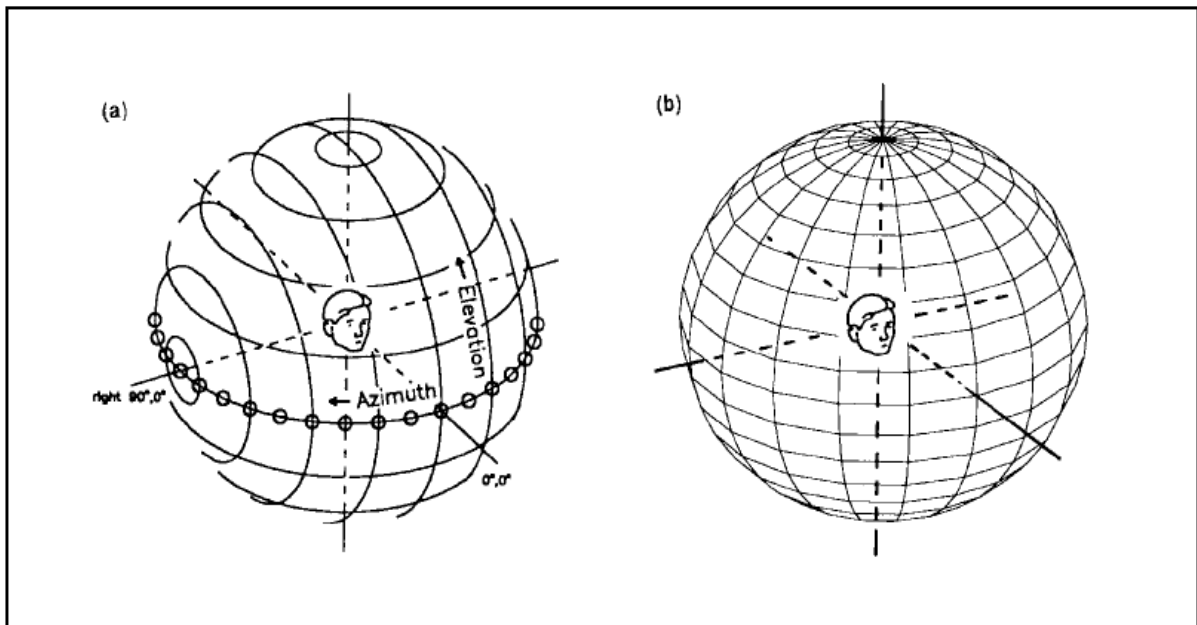


Figura 1.4.4: due diversi sistemi di coordinate: a)il sistema a polo singolo; b)il sistema a polo doppio [2].

Il vantaggio più grande di questo sistema è che è il più intuitivo, o per lo meno il sistema con cui le persone hanno più familiarità. Uno degli svantaggi è che la distanza del raggio, specificata da un particolare numero di gradi azimuth, varia come una funzione dell'elevazione. Per esempio al circolo più grande (elevazione 0°), la lunghezza dell'arco specificata in gradi è la più grande e diviene progressivamente più corta allorché ci si avvicina ai poli.

Quando si campionano funzioni HRTF a punti equidistanti nello spazio questo diviene un problema. La semplice trigonometria procurerà le accurate correzioni che si dovranno effettuare per permettere identiche

campionature di area.

Un secondo sistema di coordinate che è stato usato occasionalmente è **il sistema a polo doppio**. Questo, specifica l'elevazione allo stesso modo del sistema a polo singolo, ma definisce l'azimuth come una serie di anelli paralleli alla linea mediana e centrata ai poli ad ogni asse interaurale. Il principale vantaggio di questo sistema è che la lunghezza dell'arco azimuth è costante come funzione dell'elevazione.

Poiché il sistema a polo singolo è di gran lunga il più intuitivo, si è scelto di usare questo sistema per descrivere la localizzazione spaziale.

1.5 La testa come ostacolo acustico

Quale conseguenza della separazione delle orecchie dalla massa acusticamente opaca della testa, due differenti effetti acustici cambieranno la pressione ad ogni orecchio per una sorgente sonora localizzata lontano dal piano mediano [2].

La risultante disparità nel livello del suono ad ogni orecchio viene comunemente definita ILD (interaural level difference), comunemente ed erroneamente scambiata con IID (interaural intensity difference).

Il primo effetto che si manifesta all'orecchio che si trova sullo stesso lato della sorgente (omolaterale) del suono è dovuto alla capacità della testa di funzionare come una superficie riflettente.

Per un'onda sonora piana, a normale incidenza, la pressione del suono alla superficie di una barriera perfettamente riflettente sarà di 6 dB più alta della pressione misurata in assenza della barriera.

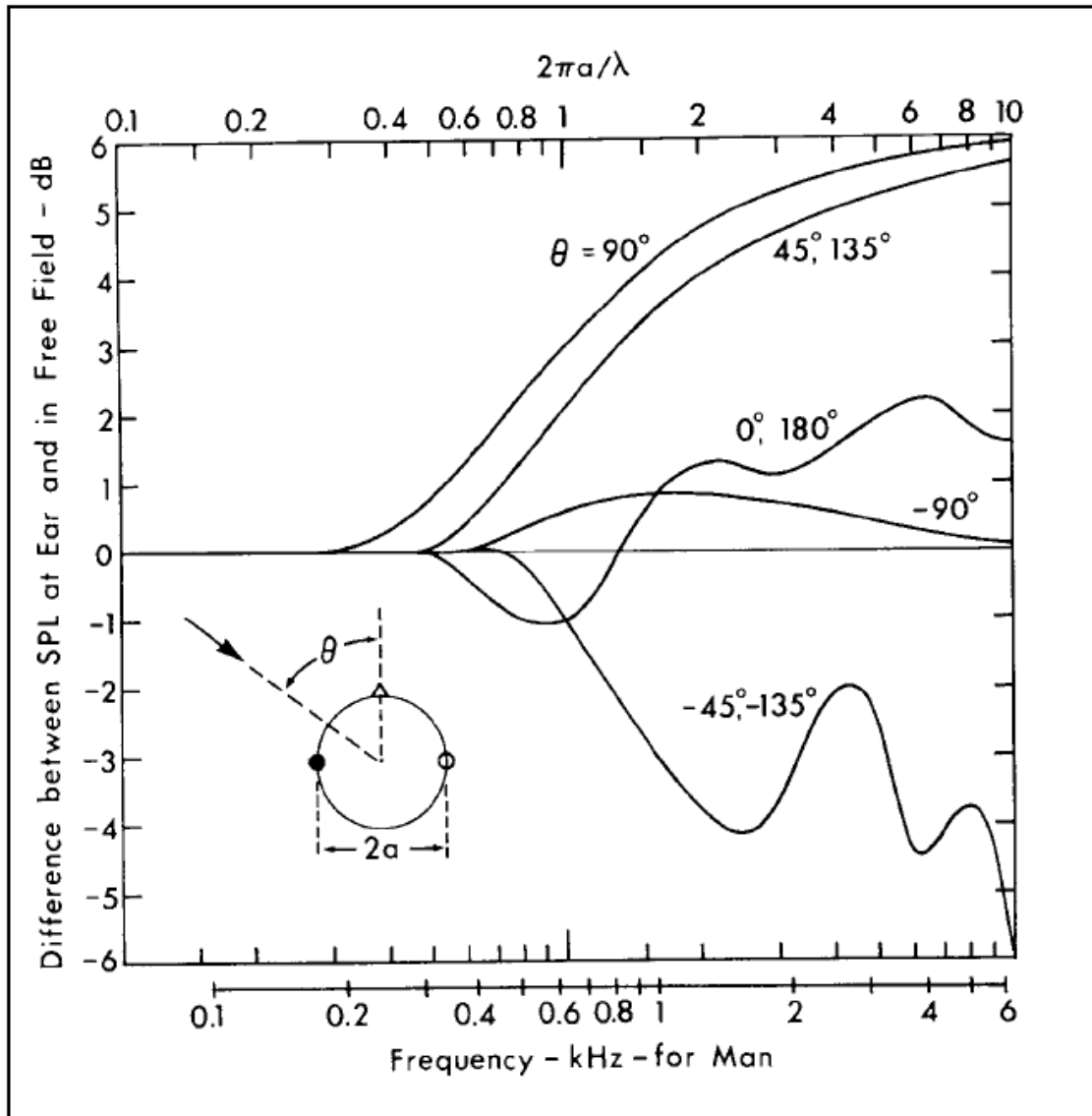


Figura 1.5: le trasformazioni del livello di pressione sonora dal campo libero rispetto alla testa sferica [2].

Così un guadagno di pressione sull'asse sarà prodotto all'orecchio omolaterale quando la lunghezza d'onda del suono è molto inferiore rispetto alla distanza interaurale.

Il secondo effetto è dovuto alla capacità della testa di diffrangere l'onda sonora.

Quando la lunghezza d'onda è dello stesso ordine di grandezza della distanza interaurale, vengono prodotti solo piccoli effetti di diffrazione.

Comunque, a lunghezze d'onda relativamente più corte, la testa funziona come un ostacolo sempre più rilevante e produce disturbi diffrattivi e riflettivi del campo sonoro.

Così per un oggetto di misura fissa quale la testa (raggio di 8.75 cm) la distribuzione della pressione del suono intorno all'oggetto dipenderà dall'angolo di incidenza e dalla frequenza dell'onda sonora piana.

La distribuzione della pressione su una sfera dura fu descritta prima da Lord Rayleigh ed in seguito sviluppata da Stewart alla svolta del secolo.

La figura 1.5 mostra i cambiamenti nel guadagno nel livello di pressione del suono (SPL: sound pressure level), relativi alla SPL in assenza della sfera, calcolati con una funzione di frequenza e l'angolo di incidenza di un'onda sonora piana.

E' da notare l'aumento asintotico fino a 6 dB per le onde a normale incidenza, dovuto al guadagno riflettivo.

In contrasto alla semplice caratterizzazione della testa come un ostacolo acustico che produce le più larghe differenze di intensità interaurale per il suono localizzato sull'asse interaurale, il modello Rayleigh-Stewart prevede che la più larga differenza interaurale avrà luogo per i suoni localizzati intorno a (+/-) 45° e a (+/-) 135° . Questo è dovuto alla natura delle interazioni diffrattive del suono che viaggia intorno alla sfera da direzioni diverse e alla loro interazione attorno all'asse dell'orecchio lontano.

Gli effetti sulla pressione del suono alla superficie della sfera prodotti dalla distanza della sorgente dal centro della sfera, sembrano essere significativi per distanze fino a $10a$, dove a è il raggio della testa, particolarmente per le basse frequenze.

Questo effetto è dovuto principalmente alla natura sferica dell'onda sonora nella vicinanza di un punto sorgente.

In sostanza, la testa funge da efficace ostacolo acustico che riflette e diffrange il campo sonoro per i suoni le cui lunghezze d'onda sono piccole in relazione alle dimensioni della testa.

Le differenze nella pressione del suono ad ogni orecchio sono in

relazione alla collocazione del suono nel libero campo ed esse, sono le più significative per le alte frequenze.

1.6 Segnali che si manifestano a causa della differenza nella lunghezza di percorso

Le differenze delle lunghezze di percorso dipendono dalla distanza e dalla collocazione angolare della sorgente rispetto alla testa [2].

La variazione nella ITD con la distanza, è realmente effettiva solamente per le collocazioni da a a $3a$; dove a è il raggio della sfera rappresentante la testa. A distanze maggiori di $3a$ il fronte d'onda è effettivamente planare.

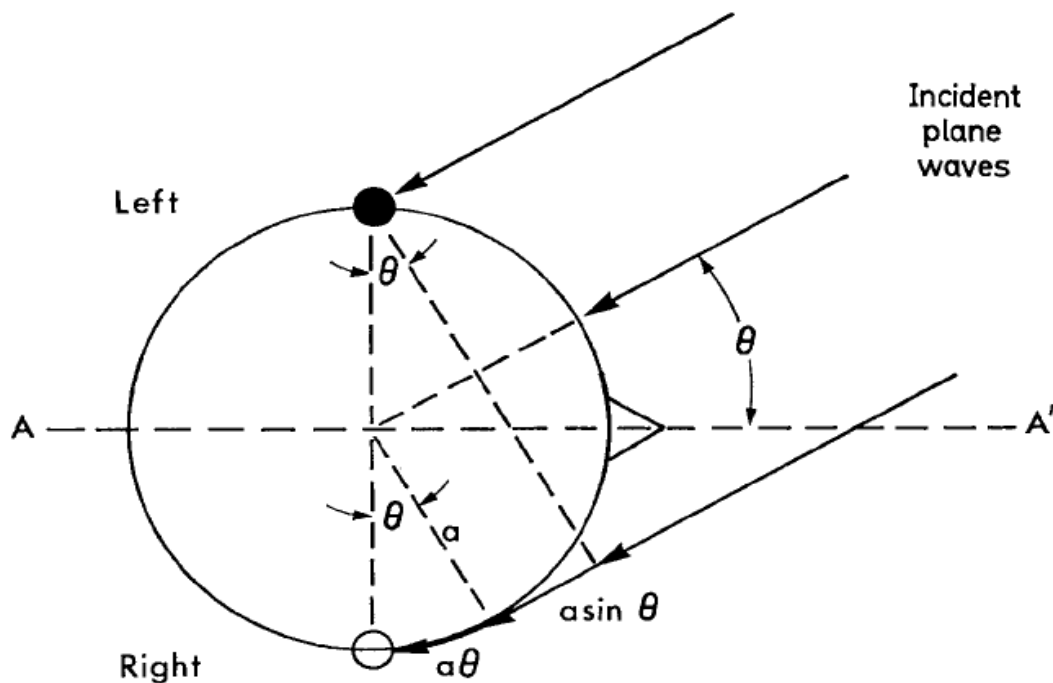


Figura 1.6: modello fisico per il calcolo della ITD [2].

La ITD prodotta dalle differenze di percorso per un'onda piana, può essere calcolata da:

$$D = a * (\theta + \sin(\theta));$$

dove D è uguale alla distanza in metri; a è il raggio della testa in metri e θ è l'angolo della fonte del suono dal piano mediano in radianti.

La differenza di tempo prodotta da questa differenza di percorso è data da:

$$\Delta t = D/c;$$

dove Δt è il tempo in secondi, c è la velocità del suono nell'aria (340 m/s).

La differenza di fase interaurale (IPD) prodotta da un segnale periodico relativamente continuo, è data da:

$$IPD = \Delta t * \omega;$$

dove ω è la velocità angolare, espressa come rapporto tra la velocità ($D/\Delta t$) ed il raggio della testa.

Per un suono continuo le differenze nella fase delle onde del suono, ad ogni orecchio, mostreranno due angoli di fase: α e $360^\circ - \alpha$.

Se questi sono segnali continui, non vi sono indicazioni a priori di quale orecchio sia predominante.

Questa informazione deve derivare dalla frequenza dell'onda sonora e dalla distanza tra le due orecchie.

Supponendo che la massima fase di differenza abbia luogo sull'asse interaurale, le uniche differenze di fase certe avranno luogo per le frequenze le cui lunghezze d'onda (λ), sono più grandi del doppio della distanza interaurale. A queste frequenze la IPD sarà sempre inferiore a 180° e quindi il segnale è chiaro.

Il contributo dato dal ritardo della risposta ad un impulso sonoro tra le due orecchie, può essere scorporato, attraverso diverse approssimazioni, dal resto delle informazioni spettrali contenute nella risposta stessa. Per una sorgente posta sul piano orizzontale, l'approssimazione più comune,

indipendente dalla frequenza del suono, è data dalla formula di Woodworth:

$$\text{ITD} = (a/c) * (\sin \theta + \theta); [0 \leq \theta \leq \pi/2]$$

$$\text{ITD} = (a/c) * (\sin \theta + \pi - \theta); [\pi/2 \leq \theta \leq \pi]$$

dove c è la velocità del suono nell'aria (circa 340 m/s).

Affinché regga l'assunzione delle onde sonore planari, la sorgente deve essere posta lontano dall'ascoltatore.

Si osserva però che in realtà l'ITD è dipendente dalla frequenza: in particolare, alle basse frequenze il ritardo è circa 1.5 volte maggiore rispetto alle frequenze più alte.

Sono dunque state introdotte ulteriori approssimazioni che tenessero conto di questa dipendenza, sia sotto forma di formule sulla falsa riga di quelle di Woodworth che di filtri nel dominio della frequenza ottenuti attraverso una ricostruzione a fase minima (tecnica attraverso cui vengono rimossi dalla fase della risposta impulsiva tutti i termini lineari, associati al puro ritardo temporale: è usata per allineare nel tempo le risposte impulsive) della risposta impulsiva.

Tuttavia, confrontando gli output della formula di Woodworth con dei risultati sperimentali, si è visto che quest'ultima approssima abbastanza bene i dati empirici.

E' stato inoltre verificato sperimentalmente che più la sorgente si allontana dal piano orizzontale, ossia più aumenta l'elevazione della sorgente, più l'ITD si riduce.

Quest'effetto, può essere incluso nella formula di Woodworth aggiungendo un termine moltiplicativo:

$$\text{ITD} = [(a/2c) (\sin \theta + \theta)] \cos \phi;$$

dove ϕ è l'angolo di elevazione della sorgente.

Infine, se si avvicina la sorgente all'ascoltatore, l'ITD aumenta di circa 150 μ s. Tuttavia, l'apparato uditivo umano appare insensibile a ritardi inferiori a 700 μ s; per questo motivo, gli effetti dell'avvicinamento della sorgente possono essere trascurati, nel modellare l'ITD.

1.8 Individuazione della distanza della sorgente

La misura dell'ITD e dell'ILD (a meno che la sorgente sonora non sia molto vicina alla testa), non permette di localizzare la distanza della sorgente, ma solo la direzione di provenienza del suono; ed essa è tanto più accurata quanto più la somiglianza spettro-temporale dei segnali che raggiungono i due orecchi è elevata (coerenza binaurale).

In condizioni anecoiche, ovvero in un ambiente progettato per massimizzare l'abbattimento delle riflessioni del suono (anecoico deriva infatti dal greco e significa "senza eco"), la percezione della distanza di una sorgente è correlata all'intensità del segnale: *l'intensità del suono è inversamente proporzionale al quadrato della distanza dalla sorgente.*

L'individuazione della distanza della sorgente sonora è più complicata rispetto all'individuazione dell'azimuth e dell'elevazione. Il parametro maggiormente utilizzato per l'individuazione di questo parametro è, come detto sopra, l'intensità con i relativi incrementi e decrementi: grazie all'esperienza, l'individuo impara ad associare tali variazioni alla posizione fisica della sorgente.

Quando la sorgente si trova lontana dal soggetto (nel "far field" ovvero approssimativamente a più di 1.5 m dal centro della testa del soggetto), allorchè la distanza della sorgente viene aumentata, *il livello di pressione diminuisce di 6 dB ogni volta che la distanza raddoppia;* si può assumere che i segnali direzionali siano indipendenti dalla distanza. Infatti, per grandi distanze, il fronte d'onda è approssimabile ad un fronte d'onda planare e quindi non è possibile ricavare i parametri utili per una corretta collocazione.

Avvicinando gradualmente la sorgente vicino al soggetto (entrando così nel "near field", ovvero a una distanza inferiore a 1.5m), il fronte dell'onda sarà curvato (il raggio della curvatura è direttamente in relazione alla distanza dalla sorgente) e si ha che l'ITD (Interaural Time Difference) resta quasi indipendente dalla distanza mentre l'ILD aumenta fortemente nell'intero spettro e in modo molto marcato alle basse frequenze.

1.8 Individuazione dell'elevazione della sorgente

Le nostre orecchie funzionano come delle “antenne acustiche”: le loro cavità risonanti sono in grado di amplificare alcune frequenze mentre portano alcuni fenomeni di interferenza che ne vanno ad attenuare delle altre. Inoltre, la loro risposta in frequenza è direzionalmente dipendente dalla posizione della sorgente.

Essendo la pinna in grado di riflettere meglio i suoni provenienti dal fronte piuttosto che da sopra, l'incavo della risposta è molto più pronunciato per sorgenti poste davanti al soggetto rispetto a quelle poste superiormente. Anche se la materia è ancora oggetto di studio, si ritiene che la pinna sia la parte del sistema uditivo che fornisce più informazioni circa l'elevazione della sorgente sonora. L'importanza della pinna nell'individuazione dell'elevazione della sorgente, porta, in modelli poco accurati o non specificatamente studiati sulla propria conformazione fisica, a un fenomeno detto front-back confusion che spinge il soggetto a collocare la sorgente sonora di fronte a sé quando essa si trova dietro, e viceversa.

La presenza di questo fenomeno, come il cono di confusione citato nel paragrafo 1.3, è dovuto al fatto che più punti nello spazio corrispondono ad identici ITD e ILD. Questa situazione si presenta tanto nelle applicazioni teoriche quanto nell'ascolto di tutti i giorni: in particolare i casi di front-back confusion sono dipendenti dal contenuto spettrale della sorgente sonora. Un espediente per individuare correttamente la posizione della sorgente anche in caso di front-back confusion è data dalle dynamic cues: infatti, dando la possibilità all'ascoltatore di muovere

la testa, essi riescono a utilizzare le informazioni derivanti dalle variazioni di ITD, ILD e movimenti dei massimi e minimi presenti nel contenuto spettrale per localizzare correttamente la sorgente, prima non precisamente collocabile.

1.9 Sensibilità psicofisica alle onde acustiche per una localizzazione del suono

1.9.1 Sensibilità alle differenze di tempo interaurale

Le più piccole differenze di tempo interaurale rilevabili sono relative al rumore che contiene basse frequenze, dove differenze appena notabili possono arrivare 6 μ s [2].

Per un tono di 1 kHz la sola variazione notevole nella fase interaurale è di 3°- 4°. Al di sopra di questa frequenza la soglia si eleva molto rapidamente cosicchè la differenza di fase non diviene più individuabile per le frequenze al di sopra di 1.5 kHz.

Per informazioni di fase non ambigue, cioè per frequenze di 1 kHz, la lunghezza d'onda (λ) è uguale a 34 cm e poichè la distanza media interaurale è uguale a 17.5 cm, il limite superiore per una precisa distinzione di fase interaurale è quasi esattamente $\lambda/2$.

Il limite superiore di frequenza per la sensibilità di fase interaurale è anche coerente con i limiti fisiologici imposti dalla fedeltà della decodifica di fase da parte del sistema nervoso uditivo.

La strategia del processo uditivo potrebbe essere determinata dalla durata del segnale che deve essere analizzato.

La percezione generata da un suono di media durata dipende dalla media delle diverse onde da cui il segnale è composto (il cosiddetto ascolto sintetico); mentre stimoli di durata più lunga tendono ad essere analizzati o incanalati in oggetti uditivi diversi.

1.9.1.1 Segnali multipli e "l'effetto precedenza"

Il sistema uditivo deve fare un'importante distinzione tra input che si riferiscono a distinti oggetti uditivi e input che si riferiscono a riflessi dovuti alle superfici vicine [2].

Le interazioni tra suoni incidenti e suoni riflessi, possono influire sul modo in cui il sistema uditivo valuta diversamente i componenti di inizio e di propagazione del suono.

Quando due stimoli simili vengono presentati da diverse localizzazioni nel campo libero, la localizzazione percepita dell'evento uditivo dipende dai tempi di arrivo dei due eventi:

- Localizzazione sommatória: un'immagine fusa in un punto tra le due sorgenti sonore
- Mascheramento del suono più tardivo: quando vi sono disparità temporali più grandi, il suono più tardivo viene mascherato fino a che la disparità di arrivo nel tempo è dell'ordine dei 5 - 40 ms. Sebbene il suono tardivo venga mascherato, esso può avere qualche effetto sulla percezione dell'immagine finale, questo è il caso in particolare della presenza di una grande disparità di livello tra il primo ed il successivo suono.

Per esempio, se il secondo suono è molto più alto, esso può completamente ignorare l'effetto precedenza.

L'effetto precedenza o l'effetto Haas è un fenomeno fisico che riguarda la percezione del suono da parte del cervello: descrive l'abilità dell'orecchio di integrare tutti i suoni che lo raggiungono entro circa 50 ms, rendendo più forte il suono e dando l'impressione che tutti questi suoni provengano dalla sorgente originaria, anche se sono presenti suoni riflessi provenienti da altre direzioni. La fusione uditiva, opera al meglio durante i primi 20-30 ms; oltre i 50-80 ms, cominciano a dominare eco separate.

L'effetto precedenza è un fenomeno binaurale: in un ambiente riverberante la nostra percezione del suono riflesso, raramente interferisce con la nostra percezione del suono primario.

Il semplice espediente di bloccare un orecchio può risultare di grande aiuto alla nostra percezione del riverbero. L'enfasi sulle componenti iniziali dei segnali che arrivano ad ogni orecchio ha ovvi vantaggi per la localizzazione del suono in ambienti riverberanti.

Come si può vedere nella figura 1.7, Haas scoprì che entro un ritardo compreso tra 5 e 35 ms, il suono proveniente dall'altoparlante ritardato, deve essere aumentato di 10 dB, rispetto a quello diretto, affinché suoni come un'eco.

In questo intervallo, infatti, le componenti riflesse che arrivano da molte direzioni vengono fuse dall'orecchio in modo che il suono risultante risulti più forte e proveniente dalla sorgente diretta.

Per ritardi dell'ordine di 50-100 ms e più, le riflessioni vengono invece avvertite come eco distinte. Si parla in questo caso di *effetto precedenza* o *effetto Haas*.

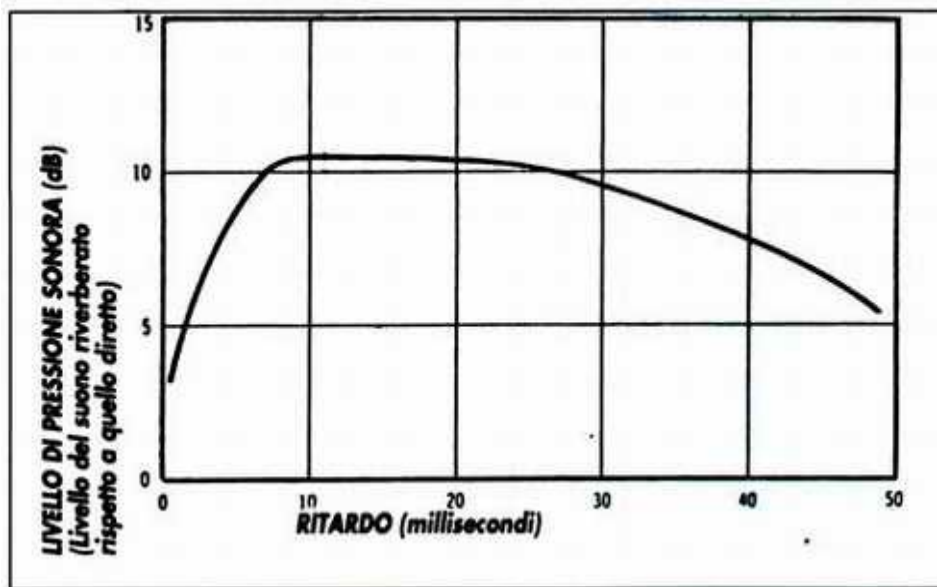


Figura 1.7: effetto Haas [1].

1.9.1.2 Misure neurofisiologiche della sensibilità relativa alla ITD

Studi su mammiferi non umani, hanno messo in evidenza che un importante prerequisito per la sensibilità interaurale alla fase di un suono, è l'abilità, che ogni orecchio ha, nel decodificare la fase di tale segnale in modo monaurale.

Dopo una prima decodifica, in un secondo processo, ci sarà una comparazione binaurale della fase di entrata ad ogni orecchio.

Questo coinvolgerà diversi meccanismi neurali che hanno specifici limiti di decodifica.

Riguardo alla decodifica monaurale delle caratteristiche di fase di un suono; i recettori uditivi dell'orecchio interno, decodificano soltanto componenti di media e bassa frequenza.

Un meccanismo che riduce l'incertezza sulla codifica delle frequenze acustiche è la proprietà di molte cellule di avere una risposta "phase-locked" allo stimolo acustico.

Molte cellule cigliate rispondono soltanto quando il segnale acustico è in una precisa fase del suo ciclo.

Tale modalità di risposta delle fibre nervose nei mammiferi decade rapidamente per le frequenze più alte di 1.5 kHz. Al di sotto di tale frequenza, il sistema uditivo è sensibile alla fase interaurale.

Esperimenti psicofisici con gli umani ed altri primati dimostrano differenze nella distinzione della soglia di 3° e 11° rispettivamente nella fase interaurale [2]. Questo meccanismo per il rilevamento di piccole differenze di tempo interaurale, fu proposto per primo da *Jeffress*:

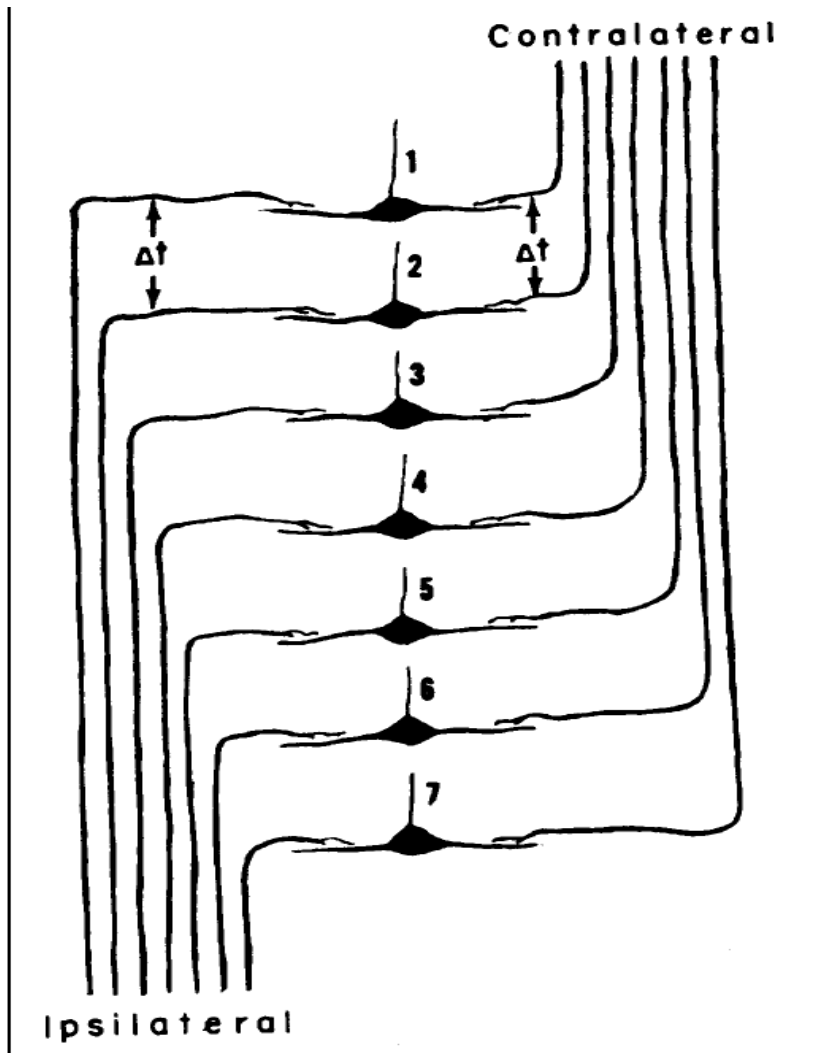


Figura 1.8: il modello di Jeffress del codice ITD [2].

Quando il suono è collocato al di fuori della linea mediana, c'è una differenza nelle lunghezze di percorso dalla sorgente ad ogni orecchio. Questo risulta in una differenza nel tempo di arrivo ad ogni orecchio. Jeffress propose che una differenza anatomica nelle lunghezze di percorso, poteva essere usata dal sistema nervoso uditivo per decodificare le ITD. Essenzialmente, le informazioni da ogni orecchio convergerebbero nel sistema nervoso lungo percorsi neuronal di diversa lunghezza. Per esempio, nel nucleo descritto in figura, vi sono una serie di neuroni che ricevono informazioni sia dall'orecchio omolaterale che da quello controlaterale. Il neurone 1 ha la lunghezza di percorso più corta dall'orecchio controlaterale e la lunghezza di percorso più lunga

dall'orecchio omolaterale. Se questo neurone rispondesse solamente e quando i segnali arrivassero ad entrambi gli orecchi, in maniera coincidente (per esempio funzionasse come un "individuatore di coincidenza"), allora il neurone 1 sarebbe selettivo per quei suoni con un grande ritardo interaurale, favorendo l'orecchio omolaterale.

Cioè, se il suono arriva prima all'orecchio omolaterale, avrà un percorso più lungo da fare fino al neurone 1; piuttosto che il suono che arriva dopo all'orecchio controlaterale. In questo modo, differenze molto piccole nelle ITD, potrebbero essere convertite in un codice di collocazione neurale dove ciascuno dei neuroni in questa serie (da 1 a 7), corrisponde ad un codice per una particolare differenza di tempo interaurale. La risoluzione di un tale sistema dipende dalle velocità di conduzione delle fibre che portano le informazioni ad ogni rivelatore di coincidenza e ad ogni reazione sinaptica del neurone rivelatore.

Gli stimoli di alta frequenza, modulati sull'ampiezza, possono essere lateralizzati sulla base dell'ITD.

1.9.2 Sensibilità alle differenze di livello interaurale

Esercitando misurazioni di pressione sonora a livello delle orecchie o all'interno dell'orecchio di soggetti sperimentali, gli studiosi hanno esaminato le onde ILD [2].

Il padiglione e altre strutture della periferia uditiva agiscono in modo da produrre un ricevitore direzionalmente selettivo, dipendente dalla frequenza.

Il sistema uditivo è sensibile a ILD dell'ordine di 1 dB o 2 dB, al di sopra dei range delle basse frequenze.

Si può eseguire anche un'analisi binaurale attraverso la frequenza, piuttosto che all'interno della frequenza.

Capitolo 2

Virtual Classroom

2.1 Analisi dello scenario di un tipico ambiente di classe

Si consideri lo scenario esemplificato in figura 1.9, in cui un insegnante insegna in un tipico ambiente di classe, dove solo alcuni alunni sono fisicamente presenti mentre il resto partecipa a distanza, per esempio da casa [3].

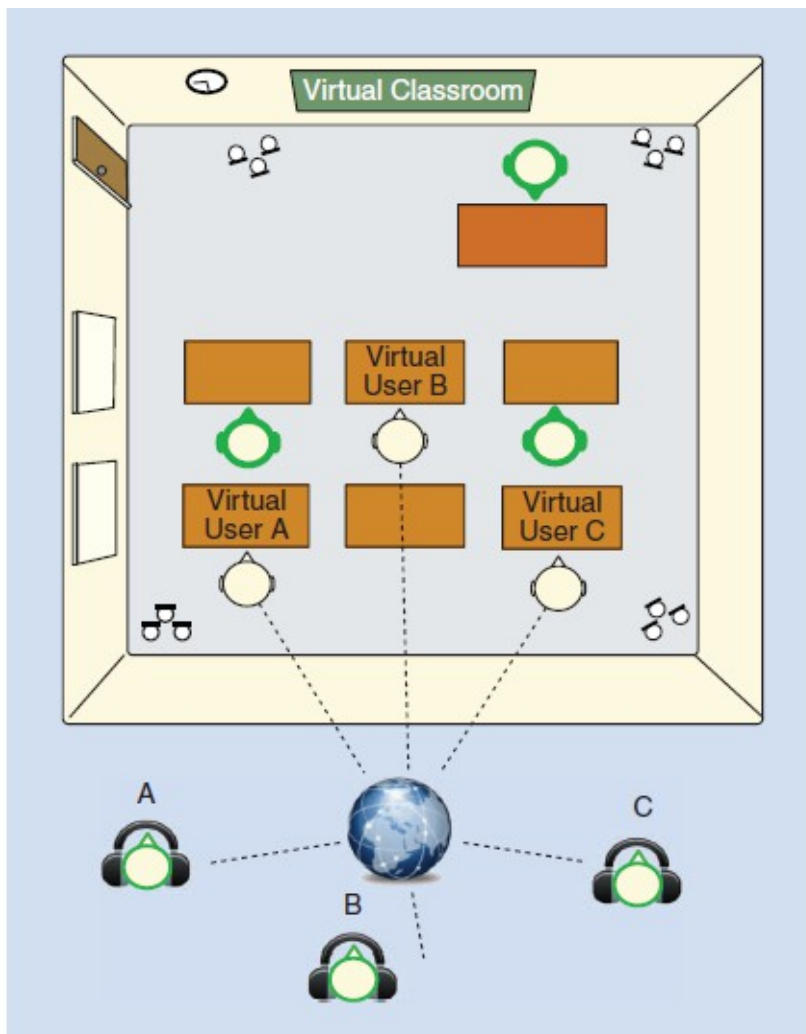


Figura 1.9: A virtual classroom scenario [3].

Come illustrato nella figura 1.9, la scena sonora è catturata usando diverse apparecchiature microfoniche distribuite, con posizioni conosciute. L'obiettivo è assistere uno studente lontano per partecipare virtualmente in una classe dalla sua posizione favorita, per esempio vicino all'insegnante, tra l'insegnante e un altro alunno coinvolto nella discussione, o al suo banco preferito; sintetizzando i segnali binaurali per l'ascoltatore virtuale desiderato (VL=virtual listener), in posizione $D(vl)$.

Questi segnali binaurali sono generati al lato riproduzione in base all'informazione audio e all'informazione sulla posizione, cosicché l'alunno potrebbe ascoltare il suono sintetizzato su delle cuffie, su un portatile o su un qualsiasi altro strumento mobile che possa far funzionare contenuti multimediali.

Il processo per raggiungere questo obiettivo consiste nel generare i segnali binaurali per la riproduzione con le cuffie; in modo che il suono risulti percettivamente simile al segnale che sarebbe stato registrato da un microfono fisico situato nella stessa posizione.

Il guadagno generale per il suono diretto nel canale può essere diviso in 3 componenti:

$$G_i(k,n) = D_s(k,n) * H_{hrtf,i}(k,n) * B(k,n,dipls)$$

Il primo guadagno $D_s(k,n)$ è un fattore che compensa per la propagazione delle onde dalla posizione $dipls$ (posizione della sorgente sonora) alla posizione dvl (posizione del virtual listener); e da $dipls$ a dl (posizione del microfono) per il segnale diretto valutato alla posizione del microfono di riferimento.

Il secondo guadagno $H_{hrtf,i}(k,n)$ è una funzione di trasferimento complessa legata alla testa (head-related transfer function (hrtf)), rispettivamente per l'orecchio sinistro o destro, dipendente dalla DOA (Direction of Arrival) del segnale, con riguardo alla posizione e alla direzione di esposizione del virtual listener (VL). Tali H_{rtf} , dipendenti dai DOAs rispetto alla posizione di ascolto virtuale, assicurano che gli indizi spaziali siano corretti.

Oltre a creare la sensazione reale di essere presente nella classe vera, la selettività spaziale dell'utente in analisi, può essere raggiunta con il terzo guadagno $B(k,n,dipls)$ che rende possibile l'amplificazione o l'attenuazione di suoni direzionali emessi da $dipls$ come desiderato.

La compensazione di propagazione e i guadagni di selettività spaziale sono fattori tipicamente reali: le fasi delle componenti dirette e diffuse, risultano infatti essere uguali a quelle osservate al microfono di riferimento.

In via di principio, ogni funzione di selettività spaziale può essere definta con il terzo guadagno.

I segnali diffusi di uscita $Y_{d,i}(k,n)$, per il canale di cuffia sinistro e destro, sono decorrelati; cosicchè la coerenza tra $Y_{d,left}(k,n)$ e $Y_{d,right}(k,n)$ corrisponde alla coerenza di fondo dell'ascolto binaurale.

2.2 Riproduzione con matlab di un ambiente di ascolto virtuale

```
function [ yout, Fs] = vocefinale( teta,d,nomefile,sec,start)
```

```
[y, Fs,nbits,opts] = wavread(nomefile);
```

```
v = 1484; [1]
```

```
rtesta = 0.0875; [2]
```

```
beta = abs(teta)/(pi/2); [3]
```

```
delta = rtesta* abs(sin (teta)); [4]
```

```
nsd = round((delta/v)*Fs);[5]
```

```
lungh=min(round(sec*Fs),length(y)); [6]
```

```
yout=zeros(lungh,2); [7]
```

```
N0=round(Fs*start); [8]
```

```
if teta >= 0
```

```
    alfad = (1/d); [9]
```

```
    alfas = (1/(d+delta))*(1+beta); [10]
```

```
    for i=nsd+1:lungh
```

```
        yout(i,1) = y(i+N0)*alfad; [11]
```

```
        yout(i,2) = y(i+N0-nsd)*alfas; [12]
```

```
    end
```

```
else alfad = (1/(d+delta))*(1+beta); [13]
```

```
    alfas = (1/d); [14]
```

```
    for i=nsd+1:lungh
```

```
        yout(i,1) = y(i+N0-nsd)*alfad; [15]
```

```
        yout(i,2) = y(i+N0)*alfas; [16]
```

```
    end
```

```
end
```

```
end
```

Linee di codice con i rispettivi commenti:

- [1] `v = 1484;` %velocità del suono nell'acqua(m/s): ipotizzando che il suono nell'attraversare la testa, incontri un fluido come l'acqua.
- [2] `rtesta = 0.0875;` %raggio della testa(m).
- [3] `beta = abs(teta)/(pi/2);` %attenuazione del segnale legata alla direzione di provenienza dello stesso(dipende dall'angolo formato dal segnale con il piano mediano).
- [4] `delta = rtesta* abs(sin (teta));` %percorso aggiuntivo per l'orecchio più lontano dovuto all'ostacolo rappresentato dalla testa.
- [5] `nsd = round((delta/v)*Fs);` %numero di campioni di ritardo (considerando l'intero più vicino a tale numero).
- [6] `lungh=min(round(sec*Fs),length(y));` %dando in ingresso alla funzione la durata di ascolto dell'audio(sec),lungh mi stabilisce il numero di campioni da analizzare con tale programma.
- [7] `yout=zeros(lungh,2);` %inizializzazione della matrice contenente i vettori relativi ai segnali che raggiungono le due orecchie.
- [8] `N0=round(Fs*start);` %il segnale parte dall'istante richiesto e non dall'inizio.
- [9] `alfad = (1/d);` %attenuazione dell'ampiezza del segnale legata all'inverso della distanza dalla sorgente(all'orecchio destro rispetto ad un segnale proveniente da destra).
- [10] `alfas = (1/(d+delta))*(1+beta);` %attenuazione dell'ampiezza del segnale legata all'inverso della distanza dalla sorgente e ad un percorso aggiuntivo rappresentato dalla testa, moltiplicato per un ulteriore fattore di attenuazione.
- [11] `yout(i,1) = y(i+N0)*alfad;` %il segnale raggiunge l'orecchio omolaterale alla sorgente solo con una certa attenuazione.
- [12] `yout(i,2) = y(i+N0-nsd)*alfas;` %il segnale raggiunge l'orecchio più lontano dalla sorgente con un certo ritardo e con una certa attenuazione.

- [13] `alfad = (1/(d+delta))*(1+beta);` %attenuazione dell'ampiezza del segnale legata all'inverso della distanza dalla sorgente e ad un percorso aggiuntivo rappresentato dalla testa, moltiplicato per un ulteriore fattore di attenuazione.
- [14] `alfas = (1/d);` %attenuazione dell'ampiezza segnale legata all'inverso della distanza dalla sorgente (all'orecchio sinistro rispetto ad un segnale proveniente da sinistra).
- [15] `yout(i,1) = y(i+N0-nsd)*alfad;` %il segnale raggiunge l'orecchio più lontano dalla sorgente con un certo ritardo e con una certa attenuazione.
- [16] `yout(i,2) = y(i+N0)*alfas;` %il segnale raggiunge l'orecchio omolaterale alla sorgente solo con una certa attenuazione.

Esempio di comandi per la riproduzione di una scena sonora a quattro voci:

```

durata=3;
start=10;
d1=1;
d2=3;
d3=2;
d4=1;
teta1=0;
teta2=pi/4;
teta3=pi/2;
teta4=-pi/6;
[yout1,Fs]=vocefinale(teta1,d1,'voce1',durata,start);
[yout2,Fs]=vocefinale(teta2,d2,'voce2',durata,start);
[yout3,Fs]=vocefinale(teta3,d3,'voce3',durata,start);
[yout4,Fs]=vocefinale(teta4,d4,'voce4',durata,start);
sound(yout1+yout2+yout3+yout4,Fs);

```

Esempio di comandi per la riproduzione di una scena sonora in cui una voce si inserisce in un discorso già avviato da un'altra voce:

```
durata=30;
start=0;
start1=5;
d1=1;
d2=3;
teta1=pi/4;
teta2=-pi/4;
[yout1,Fs]=vocefinale(teta1,d1,'voce1',durata,start);
[yout2,Fs]=vocefinale(teta2,d2,'voce3',durata,start);
[n,m]=size(yout1);
a=zeros(n,2);
ritardo=round(start1*Fs);
for i=ritardo+1:n
    a(i,1)=yout2(i-ritardo,1);
    a(i,2)=yout2(i-ritardo,2);
end
sound(yout1+a,Fs)
```

Esempio di comandi per la riproduzione di una scena sonora in cui una voce si inserisce ed il discorso già avviato da un'altra voce si interrompe:

```
durata=40;
start=0;
start1=10;
d1=1;
d2=4;
teta1=pi/6;
teta2=pi/2;
[yout1,Fs]=vocefinale(teta1,d1,'voce1',durata,start);
[yout2,Fs]=vocefinale(teta2,d2,'voce2',durata,start);
[n,m]=size(yout1);
a=zeros(n,2);
b=zeros(n,2);
ritardo= round(start1*Fs);

for i=ritardo+1:n
    a(i,1) = yout2(i-ritardo,1);
    a(i,2) = yout2(i-ritardo,2);
end

for i=1:ritardo
    b(i,1) = yout1(i,1);
    b(i,2) = yout1(i,2);
end
sound(b+a,Fs)
```

Esempio di comandi per la riproduzione di una scena sonora a quattro voci, dove ogni voce si inserisce dopo l'altra:

```
durata=40;
start=0;
start2=5;
start3=10;
start4=20;
d1=1;
d2=3;
d3=2;
d4=5;
teta1=pi/6;
teta2=pi/2;
teta3=-pi/2;
teta4=-pi/4;
[yout1,Fs]=vocefinale(teta1,d1,'voce1',durata,start);
[yout2,Fs]=vocefinale(teta2,d2,'voce2',durata,start);
[yout3,Fs]=vocefinale(teta3,d3,'voce3',durata,start);
[yout4,Fs]=vocefinale(teta4,d4,'voce4',durata,start);
[n,m]=size(yout1);
a=zeros(n,2);
b=zeros(n,2);
c=zeros(n,2);
d=zeros(n,2);
ritardo2=round(start2*Fs);
ritardo3=round(start3*Fs);
ritardo4=round(start4*Fs);

for i=1:ritardo2
    a(i,1)= yout1(i,1);
    a(i,2)= yout1(i,2);
end

for i=ritardo2+1:ritardo3
    b(i,1) = yout2(i-ritardo2,1);
    b(i,2) = yout2(i-ritardo2,2);
end
```

```
for i=ritardo3+1:ritardo4
    c(i,1) = yout3(i-ritardo3,1);
    c(i,2) = yout3(i-ritardo3,2);
end
```

```
for i=ritardo4+1:n
    d(i,1) = yout4(i-ritardo4,1);
    d(i,2) = yout4(i-ritardo4,2);
end
sound(a+b+c+d,Fs)
```


Riferimenti:

[1] F.Alton Everest, Manuale di acustica, Hoepli, 1996.

[2] S.Carlile, Virtual Auditory Space: Generation and Applications, "The Physical and Psychophysical Basis of Sound Localization", Landes Company, 1996.

[3] K. Kowalczyk et al., "Parametric Spatial Sound Processing", IEEE Signal Processing Magazine, vol. 32, n. 2, pp. 92-99, March 2015).