



UNIVERSITÀ
DEGLI STUDI DI TRIESTE



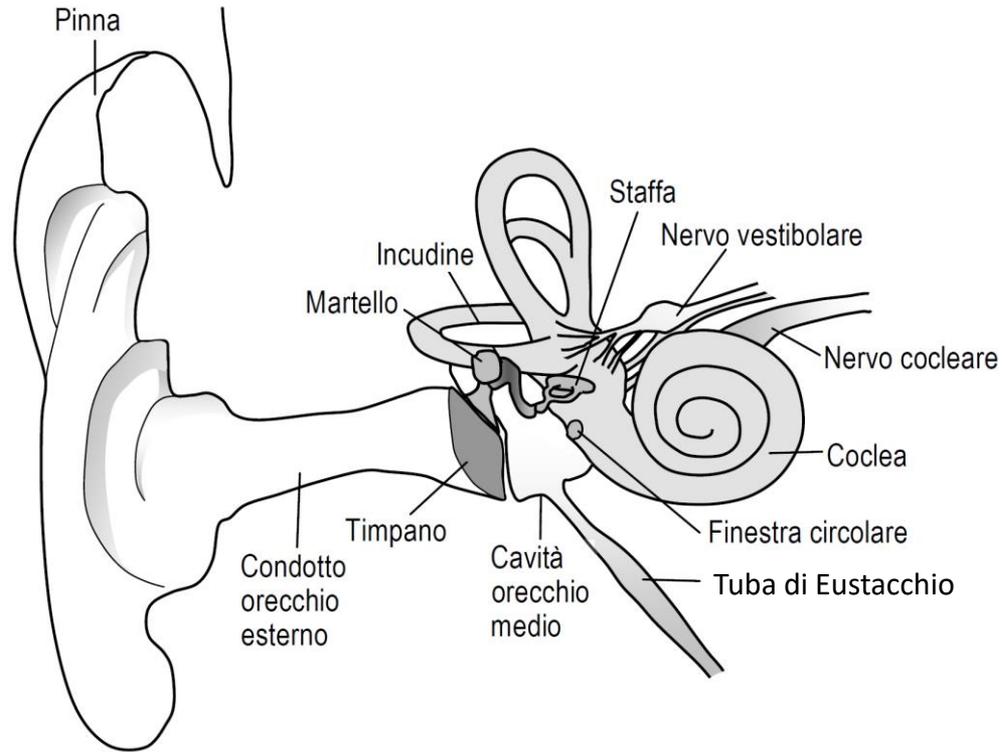
Il modello del sistema uditivo umano

A.Carini – Elettronica per l'audio e l'acustica

Il modello del sistema uditivo umano

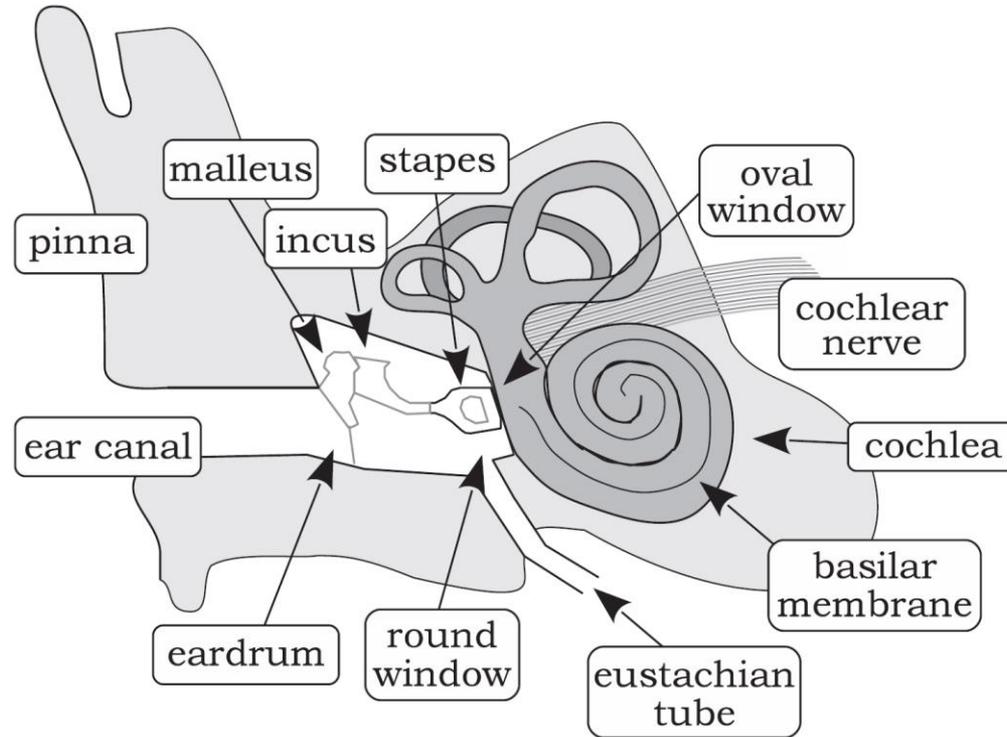
- Lo studio dell'udito umano e dei processi biomeccanici coinvolti rivela diversi passaggi non lineari.
- Quanto pensiamo di udire è molto diverso da quanto c'è nel suono fisico.
- Negli anni, la scienza ha migliorato la nostra comprensione dell'udito e l'ha sfruttata nella tecnologie, ad esempio nei sistemi di codifica.
- Esempio: la quantizzazione logaritmica con legge A o μ . Gli uomini tendono a percepire i suoni con una scala logaritmica e quindi una quantizzazione a 8bit con scala logaritmica ottiene qualità migliore di una quantizzazione lineare con 12bit.

L'organo dell'udito



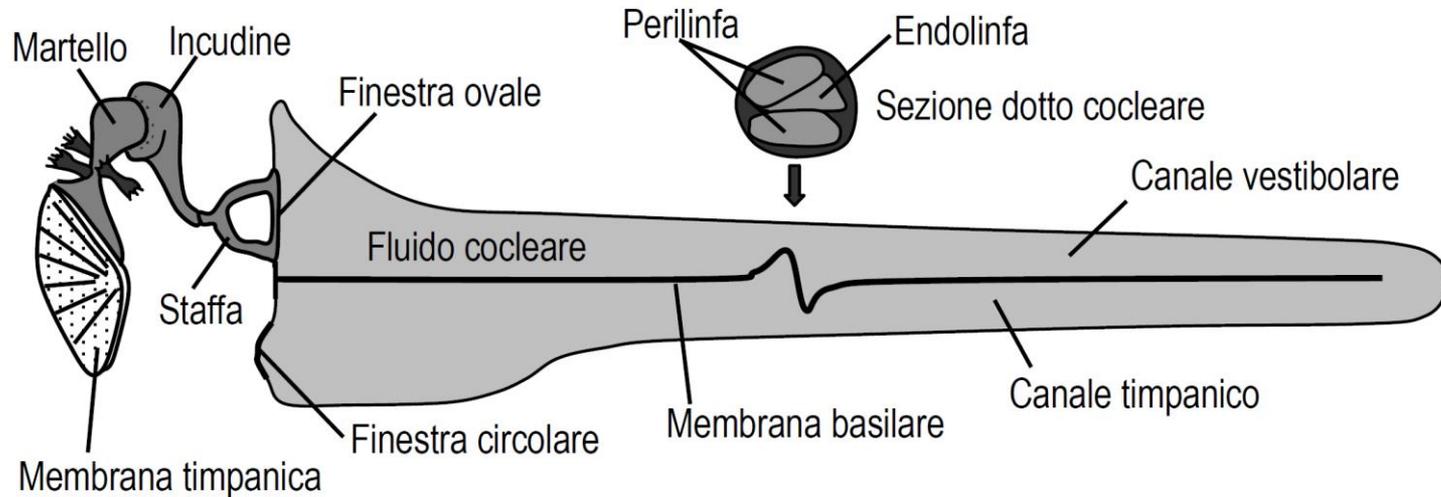
Da: A. Uncini «Audio digitale» McGraw-Hill, 2006

L'organo dell'udito



Da: S. McLoughlin «Speech and Audio Processing» Cambridge University Press, 2016

Orecchio interno

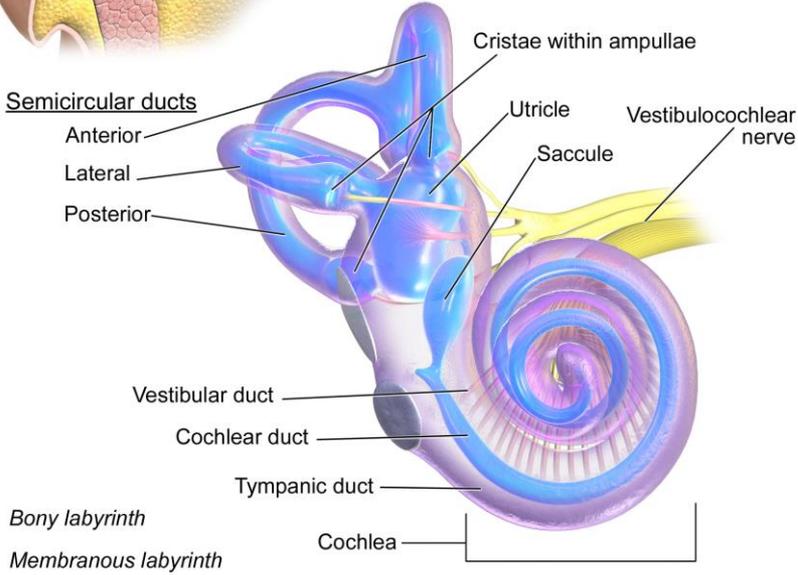
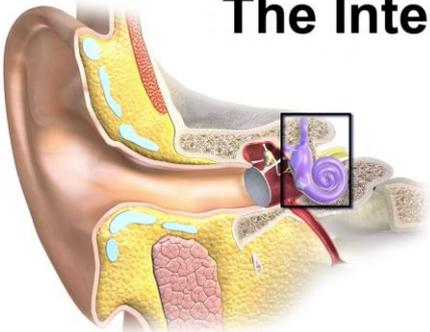


- Due membrane: basilare e di Reissner
- Tre dotti: vestibolare, medio, timpanico.
- Membrana basilare tra dotto medio e timpanico, porta gli organi di Corti, con cellule ciliate su due file (3500 interne, 13000 esterne).

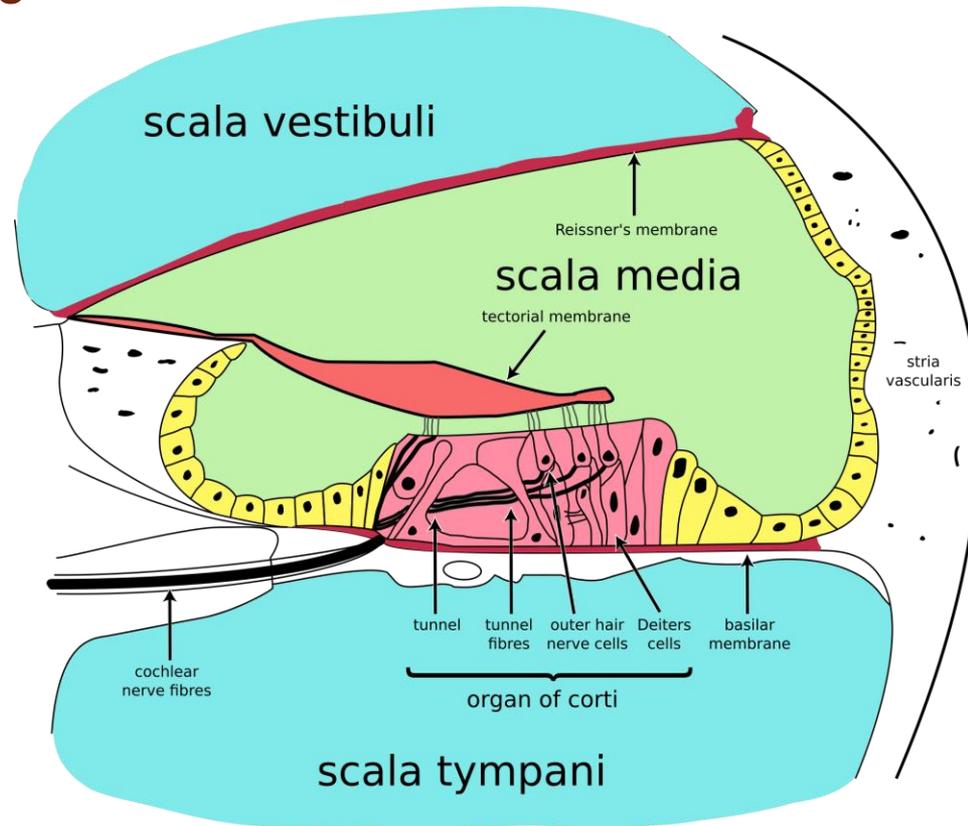
Da: A. Uncini «Audio digitale» McGraw-Hill, 2006

Orecchio interno

The Internal Ear



Orecchio interno



Psicoacustica

- Il termine «psicoacustica» deriva dalla contrazione di «psicologia dell'acustica».
- Descrive le discrepanze che esistono tra la visione puramente fisica del suono e la visione soggettiva.

Gamma di frequenza e intensità acustica

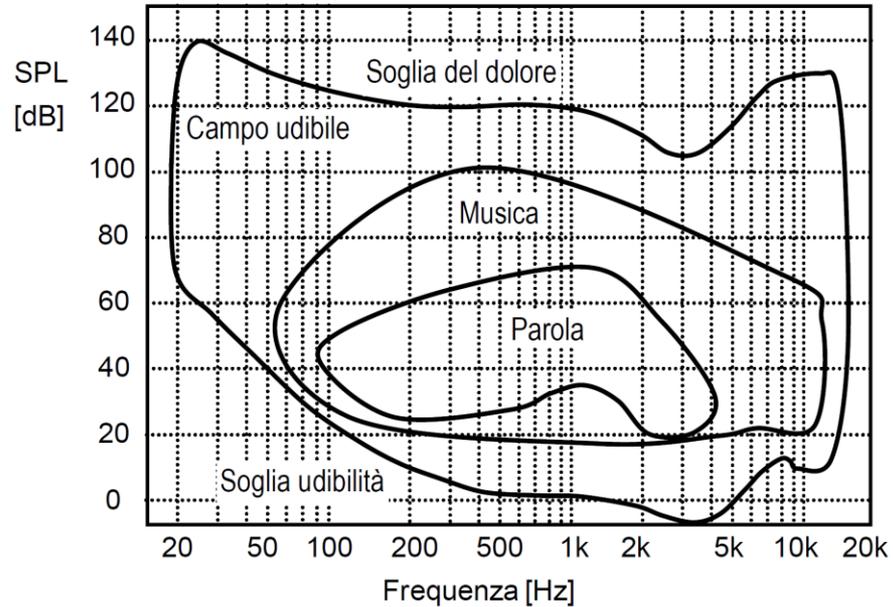


Figura 3.3 Gamma di frequenza e intensità acustica dei più comuni messaggi sonori.
Limite superiore: soglia del dolore; limite inferiore: soglia di audibilità.

Da: A. Uncini «Audio digitale» McGraw-Hill, 2006

Curve isofoniche

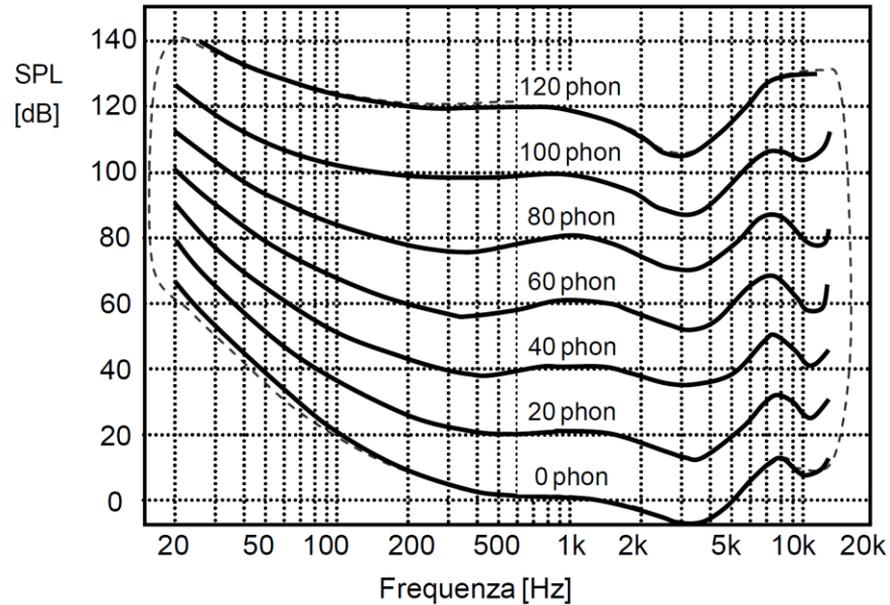


Figura 3.4 Curve isofoniche o di *loudness* di Fletcher-Munson. La famiglia di curve isofoniche prende anche il nome di *audiogramma normale* [1].

Da: A. Uncini «Audio digitale» McGraw-Hill, 2006

Curve isofoniche

```
lo= tonegen(250, 441000, 2);  
mi= tonegen(1200, 441000, 2);  
hi= tonegen(11000, 441000, 2);  
soundsc(lo, 441000);  
soundsc(mi, 441000);  
soundsc(hi, 441000);
```

Induzione di un tono a frequenza $2 f_1 - f_2$

```
fs=44100; %sample frequency
f1=tonegen(1800,fs,2);
f2=chirp([1:fs*2]/fs,2000,2,2200,'q');
left=[f1,zeros(1,fs),f1]; %repeat sounds twice
right=[f2,zeros(1,fs),f2];
```

```
soundsc([left;],fs)
%wait for this to finish, then
soundsc([;right],fs)
```

```
soundsc([left;right],fs)
```

Dovrebbe essere udibile un tono con frequenza decrescente dovuto a $2 f_1 - f_2$
 $f_1 = 1800\text{Hz}$, $f_2 = 2000\text{Hz}$, $2 f_1 - f_2 = 1600\text{Hz}$; $f_1 = 1800\text{Hz}$, $f_2 = 2200\text{Hz}$, $2 f_1 - f_2 = 1400\text{Hz}$;

Missing fundamental

```
t1=tonegen(196, 8000, 2);  
t2=tonegen(196*2, 8000, 2);  
t3=tonegen(196*3, 8000, 2);  
t4=tonegen(196*4, 8000, 2);
```

```
soundsc(t1, 8000);  
soundsc(t1+t2+t3+t4, 8000);  
soundsc(t2+t3+t4, 8000);
```

Missing fundamental

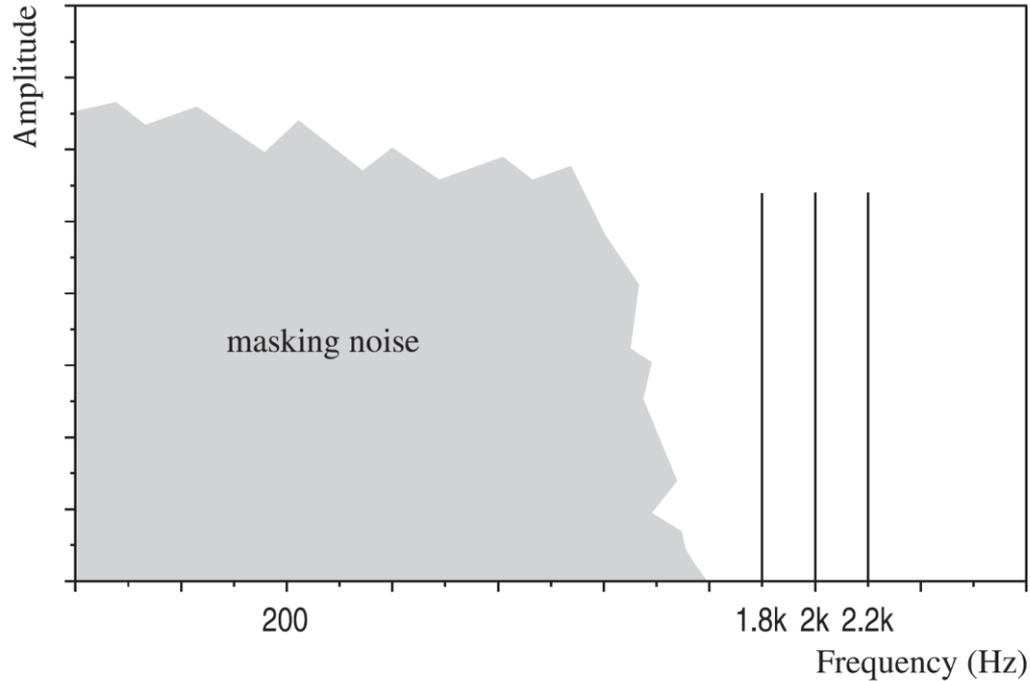


Figure 4.3 An artificial arrangement causing 200 Hz to be heard, but where the 200 Hz tone, if present, would not be discernible due to heavy masking noise.

Aggancio di fase

- L'eccitazione neurale delle cellule ciliate avviene in corrispondenza alle rarefazioni delle onde sonore, con una fase approssimativamente fissa.
- Le cellule generano i loro impulsi con una corretta frequenza e fase.
- L'orecchio si «accorda» su una certa frequenza e richiede un certo tempo per recuperare la sua funzionalità.
- Se nel periodo arriva una altra rarefazione, questa potrebbe essere persa, con il risultato che un suono più intenso potrebbe essere necessario per rendere udibile un secondo tono.
- Questo meccanismo di aggancio di fase potrebbe essere parte del fenomeno del mascheramento.

Integrazione temporale

- La risposta dell'orecchio ai toni è fortemente nonlineare.
- Per durate inferiori ai 200ms, l'intensità richiesta perché un tono sia udibile aumenta se la durata si riduce.
- Toni di durata superiore ai 500ms sono sempre udibili.
- Periodi di silenzio introdotti in toni costanti saranno rilevati a seconda della loro durata e comunque sentiti oltre i 200ms.

Post-stimulatory audio fatigue

- Dopo un suono forte in modo anomalo, la risposta dell'orecchio è ridotta durante un periodo di recupero. Si parla di *temporary threshold shift (TTS)*.
- Il grado dipende dall'intensità, dalla durata e dalla frequenza del suono anomalo.
- Con suoni a larga banda, TTS principalmente tra i 4 e 6 KHz. Inizia immediatamente e dura anche 16 ore.
- Toni puri di ampiezza superiore ai 110-120 dB SPL possono causare perdite di udito permanenti. TTS particolarmente evidente con 90-100 dB SPL.

Altezza o *pitch* di un suono

- L'altezza o pitch è una sensazione soggettiva.
- L'American Standard Association definisce il pitch come
«un attributo della sensazione uditiva in cui un suono può essere ordinato in una scala musicale»
- Il pitch non sempre corrisponde esattamente alla frequenza oggettiva del segnale. L'intensità ha un effetto sul pitch percepito alle alte o basse frequenze.
- Esiste una discordanza, particolarmente evidente sopra 1kHz tra frequenza reale e frequenza percepita.
- Quando la pressione acustica cresce, l'altezza di un suono in bassa frequenza diminuisce, ad alta frequenza aumenta.
- La scala mel (da melody) misura l'altezza percepita da un ascoltatore.

Altezza o *pitch* di un suono

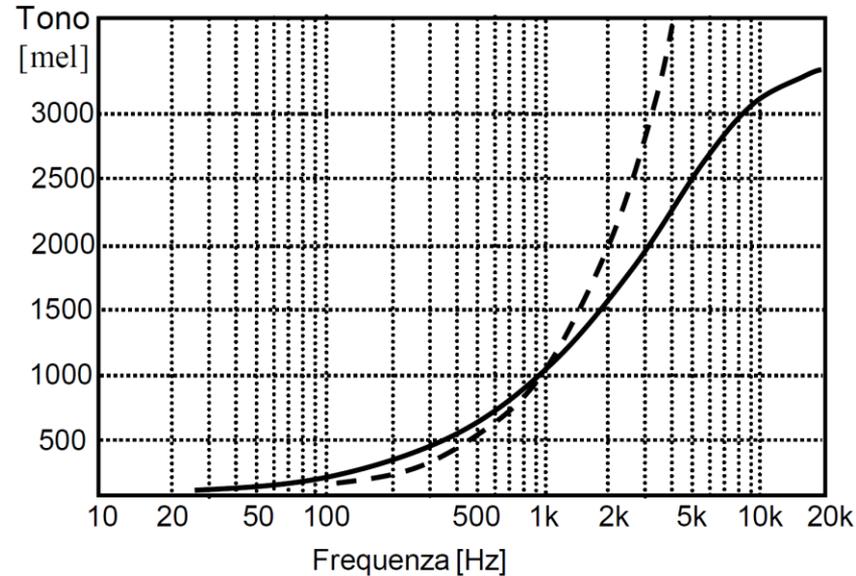


Figura 3.6 Giudizio medio di tono in [mel] vs. frequenza effettiva dell'oscillazione [Hz]. Linea continua $I = 60$ dB (curva mel). Linea tratteggiata $I > 60$ dB.

Da: A. Uncini «Audio digitale» McGraw-Hill, 2006

Le bande critiche

- Le cellule dell'organo di Corti lavorano in gruppi di ~ 1300 , ognuno dei quali occupa fisicamente circa $\sim 1.3\text{mm}$ di membrana basilare e copre una frequenza di $\sim 1/3$ di ottava.
- Ognuno di questi gruppi costituisce una banda critica.
- Quando due frequenze sono abbastanza vicine da stimolare lo stesso gruppo di cellule, e quindi cadono nella stessa banda critica, la loro distinzione risulta difficile.
- Consideriamo un semplice esperimento: sommiamo due segnali sinusoidali con frequenze tra di loro vicine:
- Al crescere della differenza tra le frequenze si ha prima battimento, poi suono aspro dissonante, e infine due suoni distinti non sgradevoli o consonanti.

$$\sin \omega t + \sin(\omega + \Delta\omega)t = 2 \cos(\Delta\omega/2)t \cos(\omega + \Delta\omega/2)t.$$

Le bande critiche

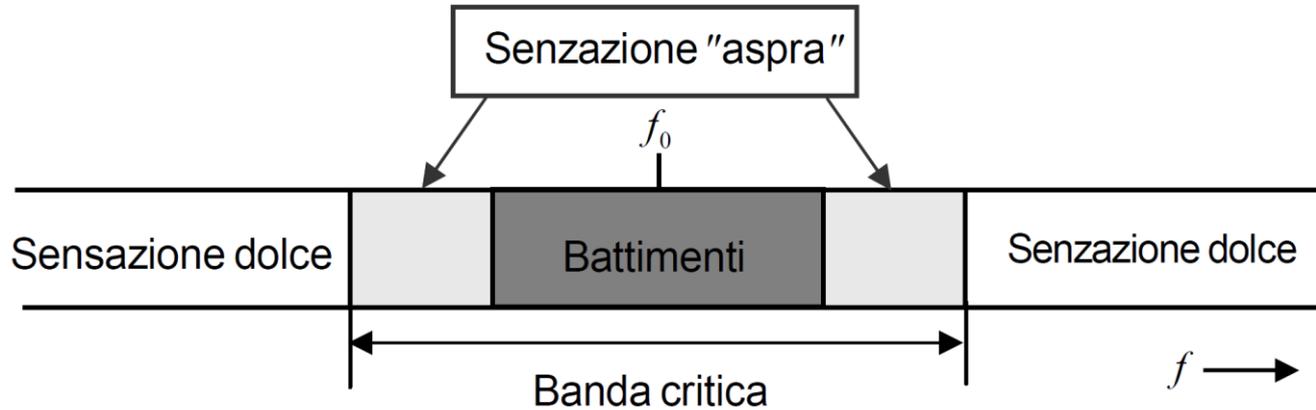


Figura 3.7 Fenomeno dei battimenti e banda critica.

Le bande critiche

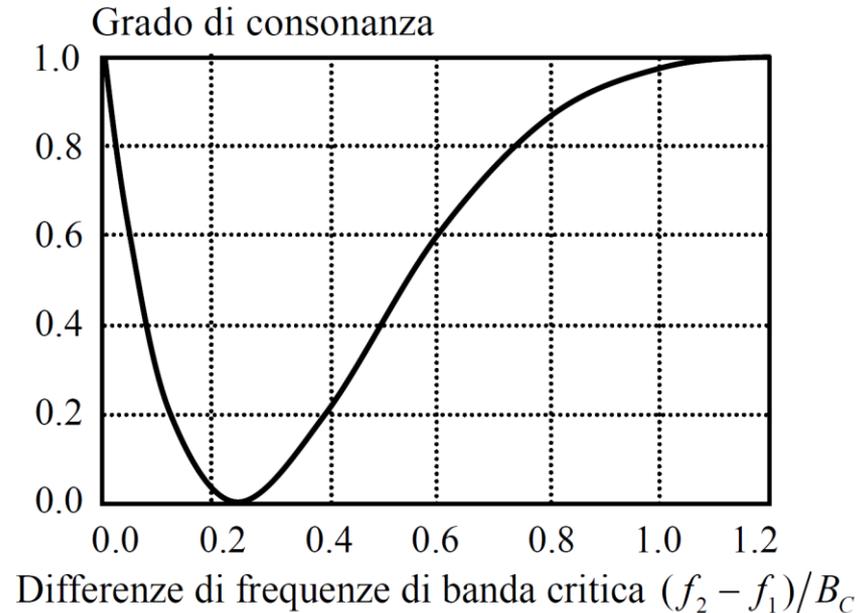
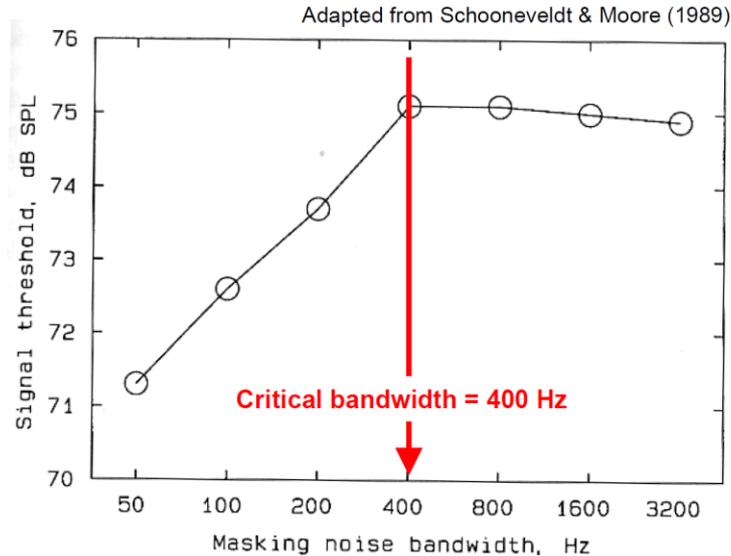
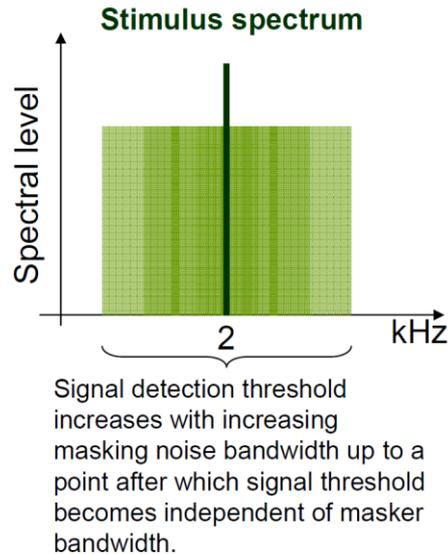


Figura 3.8 Curva di consonanza di due suoni puri (Misure di Plomp and Levelt (1965)).

The “critical band”



Schooneveldt GP, Moore BCJ. (1989). Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands. *J Acoust Soc Am.* 85(1):262-272.

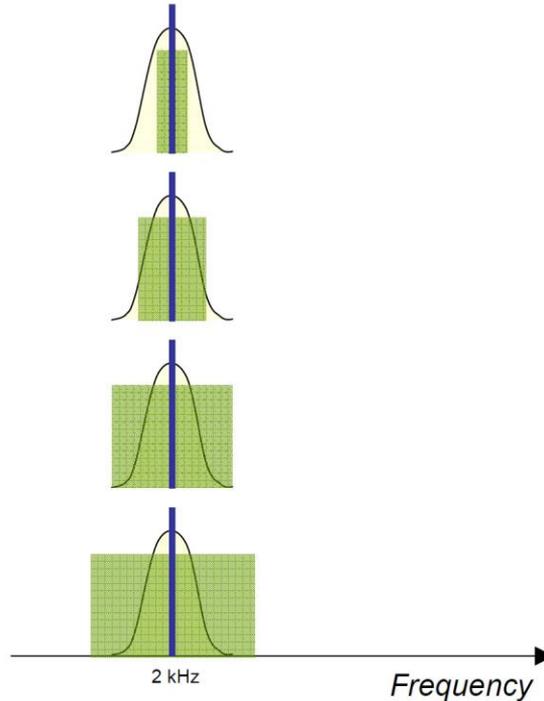
Enrique A. Lopez-Poveda, Ph.D.

An explanation of the critical band

The amount of masking increases with increasing the noise (masker) energy that gets through the filter.

Up to a point...!

Further increases in noise bandwidth do not increase the masker energy through the filter.



Enrique A. Lopez-Poveda, Ph.D.

Le bande critiche

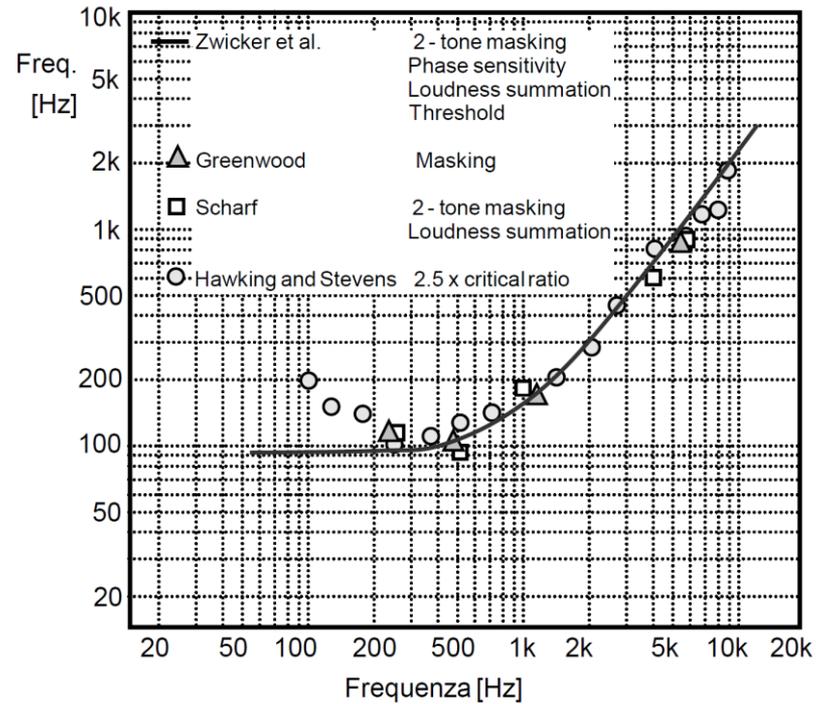


Figura 3.9 Larghezza delle bande critiche soggettive in funzione delle frequenza [5].

Le bande critiche

Table 4.4 Critical bands and corresponding centre frequencies.

Critical band (Bark)	Lower cutoff frequency (Hz)
1	100
2	204
3	313
4	430
5	560
6	705
7	870
8	1059
9	1278
10	1532
11	1828
12	2176
13	2584
14	3065
15	3630

La scala Bark

- Il Bark (da Barkhausen) è una scala di frequenze in cui la variazione di una unità (1 Bark) corrisponde a una variazione di una banda critica.
- Formula approssimata di Hermansky:

$$\Omega(\omega) = 6 \log \left(\omega/1200\pi + \sqrt{(\omega/1200\pi)^2 + 1} \right)$$

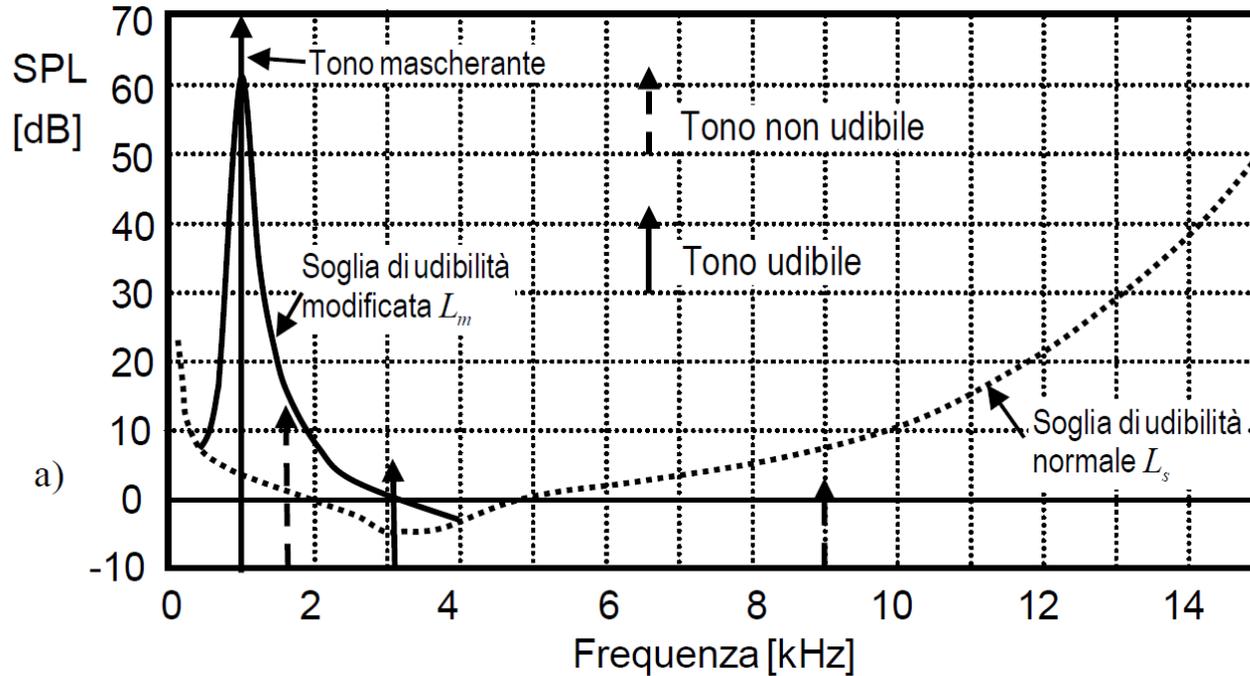
```
function [bark]=f2bark(hz)
    cn=2*hz/1200;
    bark=6*log(cn+(cn^2+1)^0.5);
```

```
function [hz]=bark2f(bark)
    hz=600*sinh(bark/6);
```

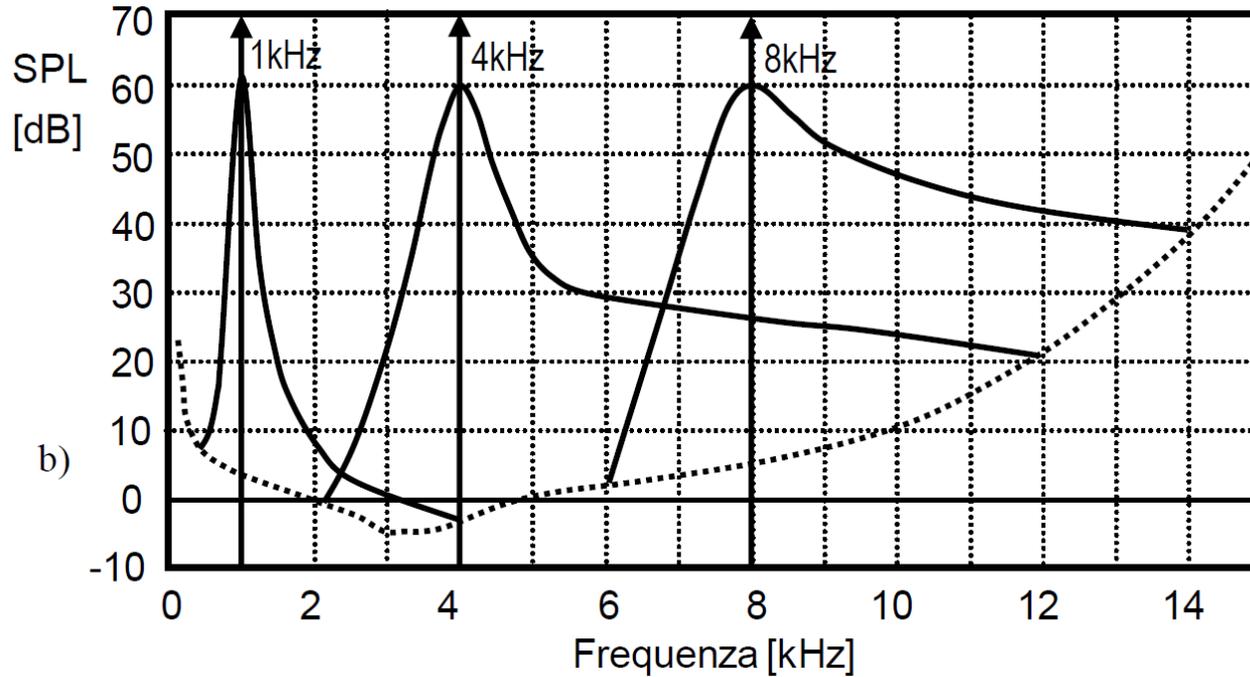
Mascheramento/Masking

- Secondo la American Standard Association:
«processo per cui la soglia di udibilità di un suono è aumentata dalla presenza di un altro suono»
«il quantitativo per cui la soglia di udibilità del suono è aumentata dalla presenza di una altro suono».

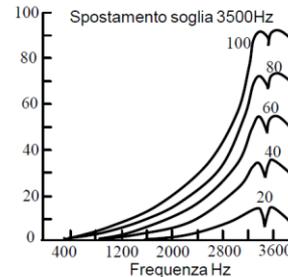
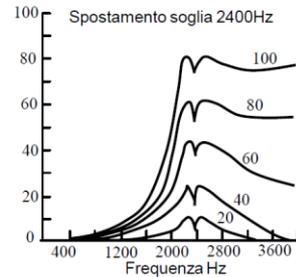
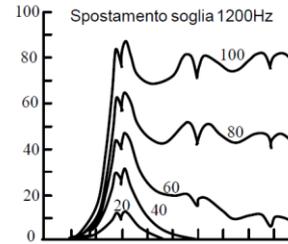
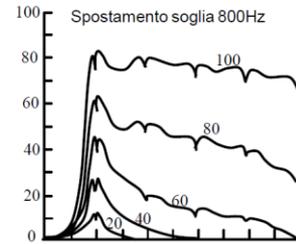
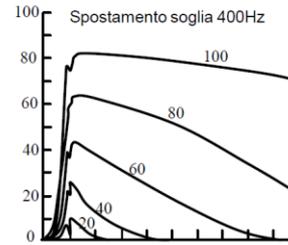
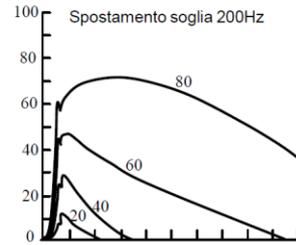
Mascheramento / Simultaneous Masking



Mascheramento / Masking



Mascheramento con toni



Mascheramento con toni e rumore bianco

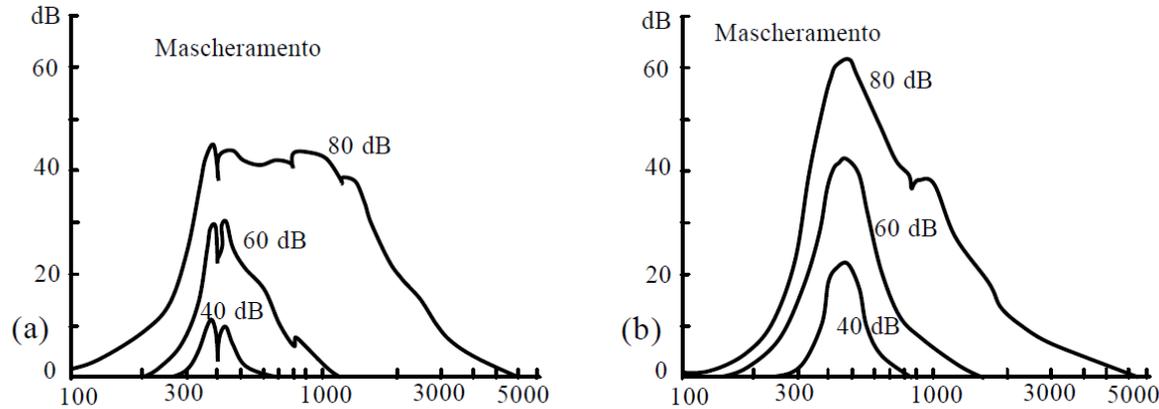


Figura 3.13 Effetto di mascheramento: a) per suono mascherante puro a 400 Hz ; b) rumore bianco tra 365 Hz e 455 Hz.

Non simultaneous masking

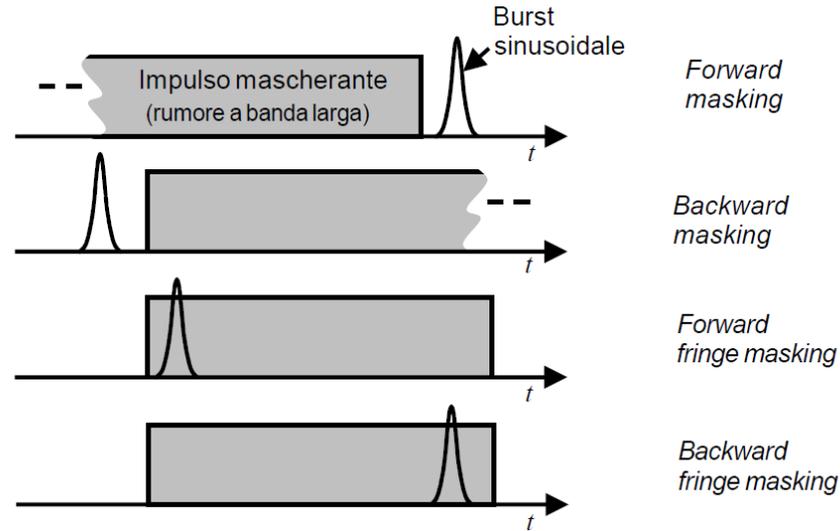


Figura 3.17 Successione temporale dei segnali da somministrare ai soggetti per lo studio dinamico del fenomeno di mascheramento.

Non simultaneous masking

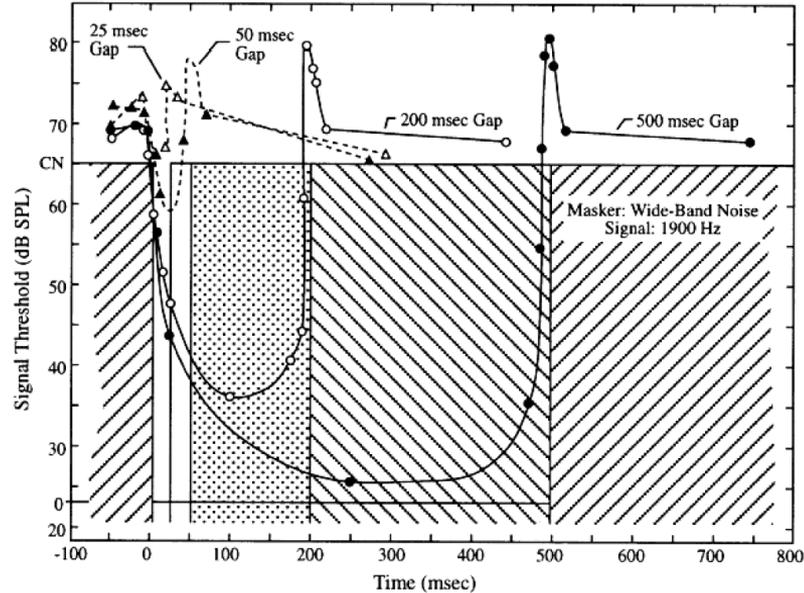


Figura 3.18 Fenomeno del mascheramento dinamico con finestra di silenzio di 25, 50, 200, 500 msec. La figura illustra simultaneamente il: *forward*, *backward*, *forward fringe* e *backward fringe masking* [12].

Discriminazione delle frequenze

- Dipende dal valore della frequenza e dall'ampiezza.
- Una figura approssimata è di circa 2Hz a 65 dB SPL per un segnale a 1 kHz.
- Dipende dalla percezione del pitch, che diminuisce con l'aumento dell'ampiezza sotto 2 kHz, aumenta con l'ampiezza sopra i 4 kHz.
- Pitch sopra i 5 kHz non possono essere discriminati (non possono nemmeno evocare la sensazione della melodia).



Precedence effect o effetto di Haas.

- Se due versioni dello stesso segnale raggiungono l'orecchio ritardate nel tempo, l'osservatore udirà la prima versione ma non la seconda e sentirà il suono provenire dalla prima sorgente.
- Questo effetto agisce con segnali che abbiano un ritardo inferiore a 50ms, anche quando il secondo segnale abbia una ampiezza 10dB superiore al primo.
- Viene spesso sfruttato per l'amplificazione di segnali in sistemi di «public addressing»
- Echi con ritardi superiori ai 65ms (corrispondenti a una differenza di cammino di 20m) vengono percepiti come distinti.
- Echi con ritardi superiori ai 100ms sono molto disturbanti.

Modelli posizionali 2D – teoria duplex

- Si basa su due grandezze:
- *Interaural time difference (ITD)*: differenza temporale con la quale la forma d'onda del suono giunge alle due orecchie.
- *Interaural level difference (ILD)*: differenza di intensità, in dB, percepita dalle due orecchie.
- Entrambe le grandezze sono importanti per la discriminazione della provenienza di un suono sul piano azimutale.

Modelli posizionali 2D – teoria duplex

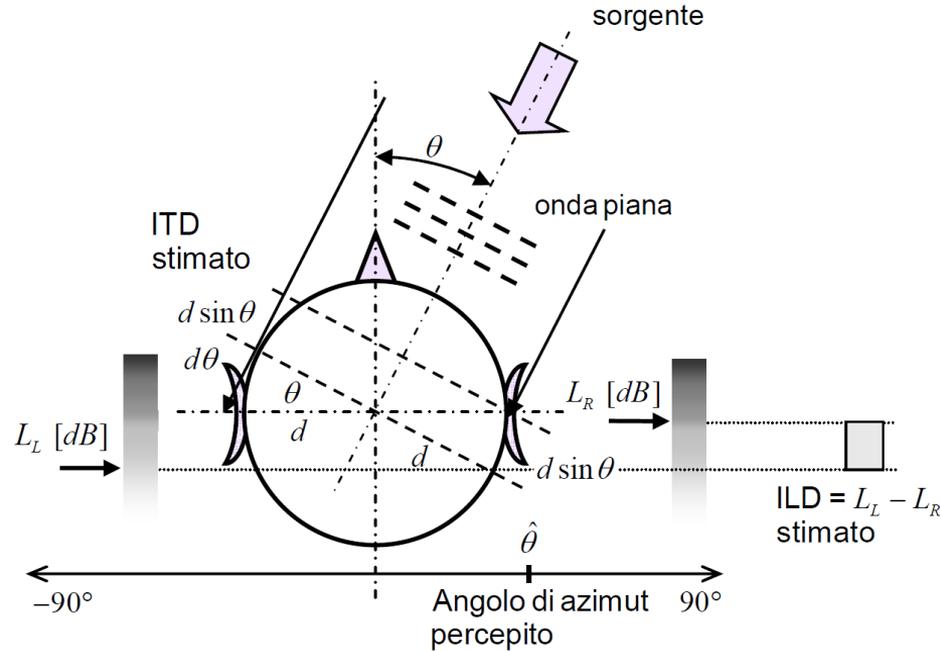


Figura 3.20 Geometria del modello Duplex. Differenza di intensità interaurale (ILD) e differenza di tempo di arrivo (ITD). Nel modello Duplex la direzione dall'angolo azimutale è valutata per mezzo di tali quantità.

Modelli posizionali 2D – teoria duplex

- Per frequenze < 1.5 kHz e onda incidente piana vale la formula di Woodworth:

$$\text{ITD} = \frac{d}{c} (\sin \theta + \theta) \cos \phi; \quad -90^\circ \leq \theta \leq 90^\circ;$$

- Con c velocità del suono, d raggio della sfera, θ angolo azimutale, ϕ l'angolo di elevazione.

Modelli posizionali 2D – teoria duplex

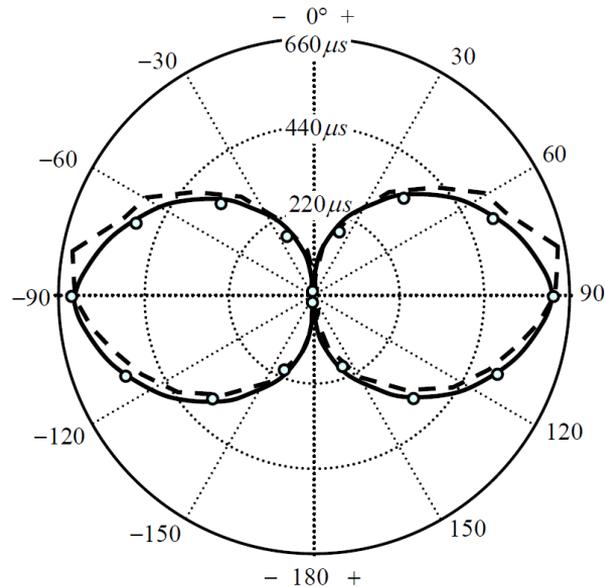


Figura 3.21 ITD sul piano orizzontale (in μs). Il tratto continuo è relativo alla equazione (3.10). I punti rappresentano una misura media valutata su 70 soggetti. La linea tratteggiata è relativa a un modello ellissoidale (tratta da [35]).

Modelli posizionali 2D – teoria duplex

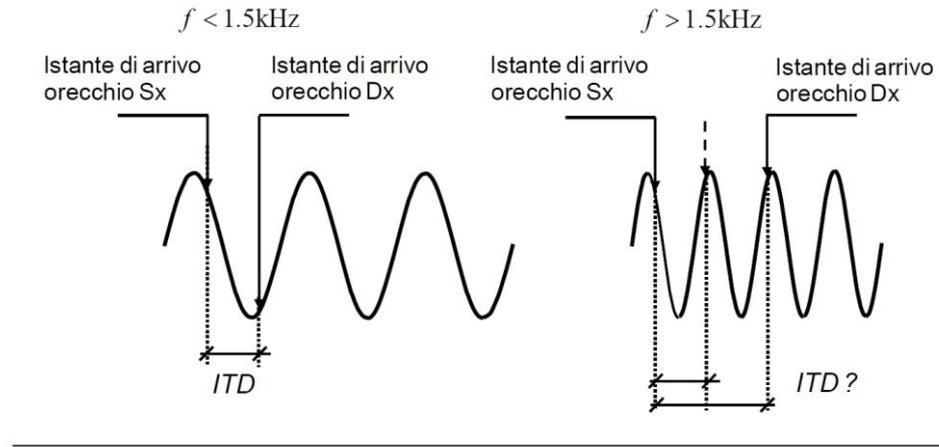


Figura 3.22 Ambiguità della ITD per frequenze > 1500 Hz (*aliasing*).

- Per $f > 1.5$ kHz diviene fondamentale l'ILD
- La relazione che lega percezione della direzione a ILD è nonlineare e legata alla frequenza. Approssimativamente, l'azimut varia in modo lineare con il logaritmo dell'ILD.

Teoria duplex – cono di confusione

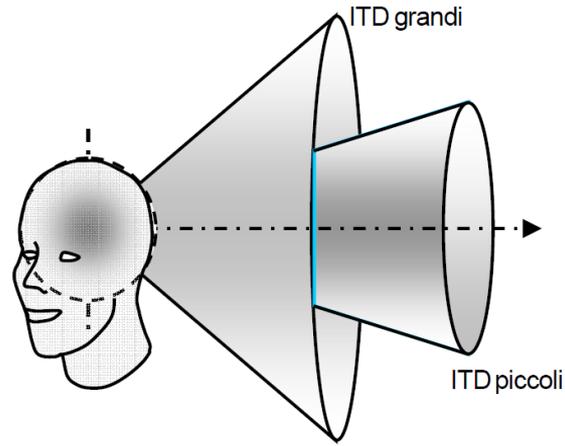


Figura 3.23 Cono di confusione calcolato considerando la testa di forma sferica e in assenza di pinna. Le sorgenti acustiche nel cono di confusione hanno la ITD e ILD identici.

Modelli posizionali 3D – Head related transfer function (HRTF)

- Per una sorgente acustica localizzata nello spazio sul punto relativo all'ascoltatore di coordinate (r, θ, φ) , si definisce *HRTF* $H_{r, \theta, \varphi}(\omega)$ il rapporto tra il livello di pressione sonora in un punto del canale uditivo in prossimità del timpano e l' SPL alla locazione della sorgente in campo libero (misurata in camera anecoica).
- Sono funzioni di 4 variabili ma siccome la misura viene fatta in campo lontano le variabili si riducono a 3: $H_{\theta, \varphi}(\omega)$

Modelli posizionali 3D – Head related transfer function (HRTF)

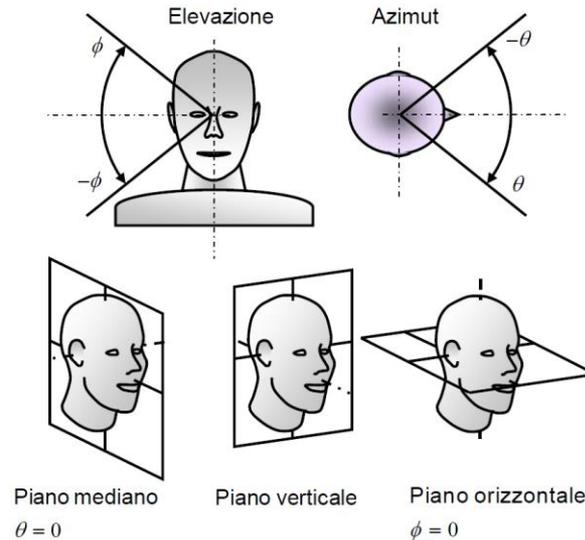


Figura 3.24 Definizione di elevazione, azimuth e dei piani di riferimento. Si assume che le coordinate relative all'orecchio destro siano $(\theta, \phi) = (+90^\circ, 0^\circ)$, mentre per l'orecchio sinistro $(\theta, \phi) = (-90^\circ, 0^\circ)$.

Modelli posizionali 3D – Head related transfer function (HRTF)

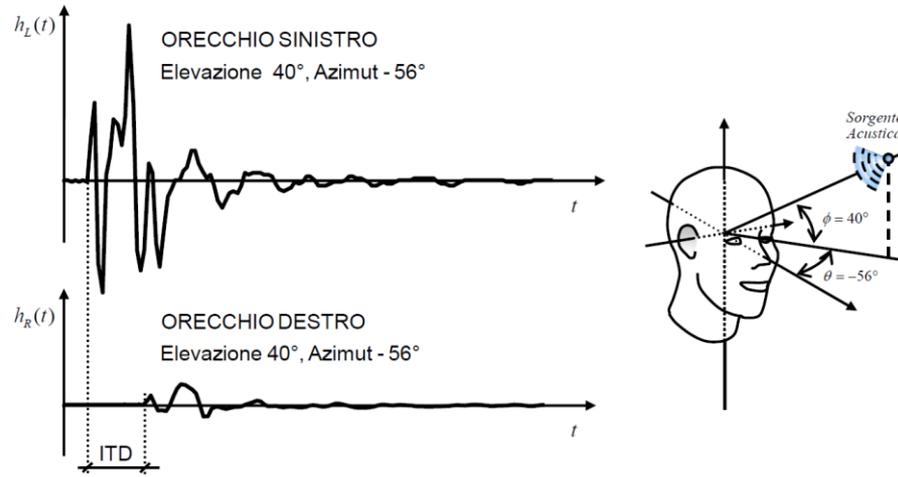


Figura 3.25 Coppia di HRIR (orecchio sinistro e orecchio destro), misurata per una posizione $(\theta, \phi) = (-56^\circ, 40^\circ)$. Si notino i valori nulli dei primi campioni relativi alla distanza della sorgente da cui è possibile stimare la ITD. (Una stima più consistente andrebbe fatta considerando la correlazione tra le due HRIR).

Modelli posizionali 3D – Head related transfer function (HRTF)

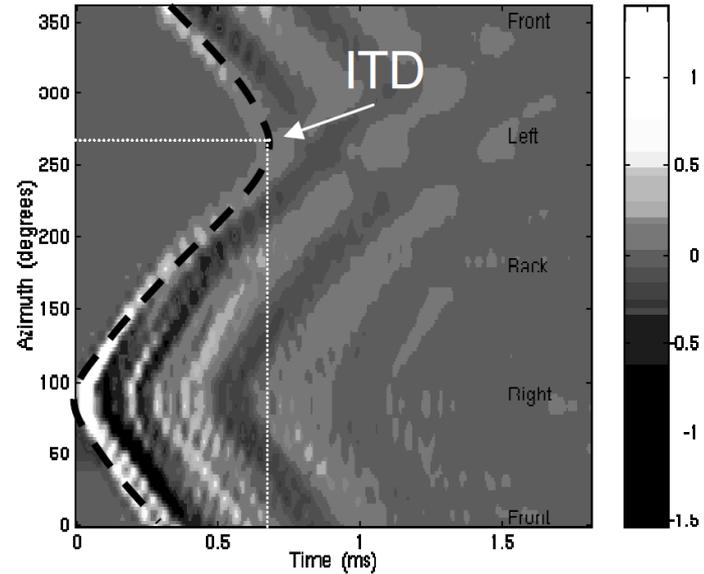


Figura 3.26 Immagine rappresentante le HRIR per l'orecchio destro valutate sul piano azimutale, misurate per una posizione $\theta = (0^\circ, 360^\circ)$.

Modelli posizionali 3D – Head related transfer function (HRTF)

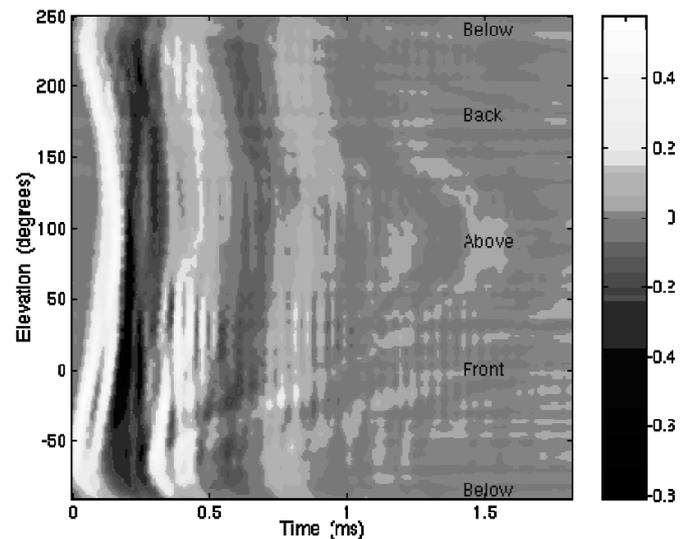


Figura 3.27 Immagine rappresentante le HRIR per l'orecchio destro valutate sul piano mediano ($\theta=0^\circ$), misurate per una posizione $\phi = (-100^\circ, 250^\circ)$.

Modelli posizionali 3D – Head related transfer function (HRTF)

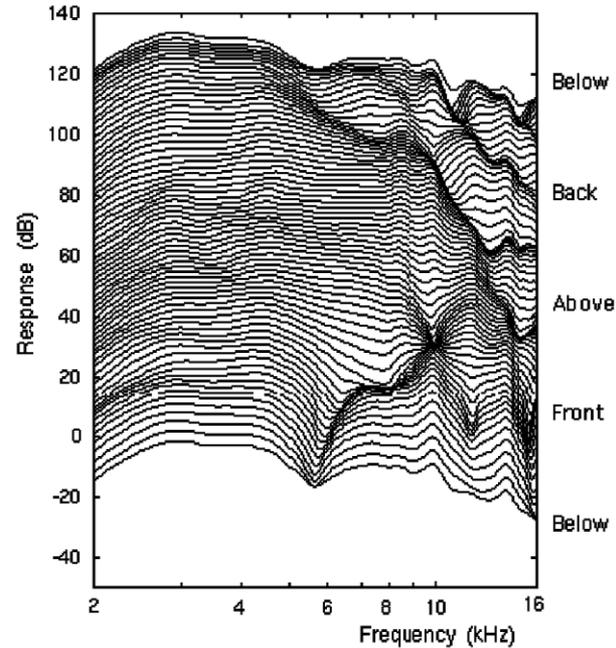
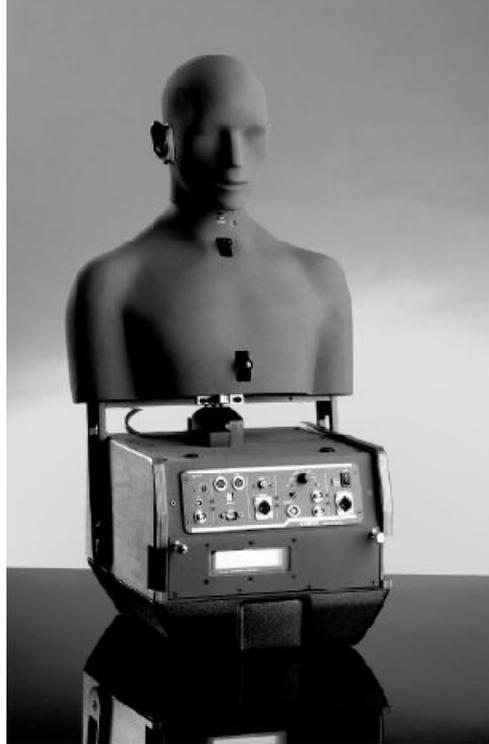


Figura 3.29 Risposte in ampiezza delle HRTF valutate sul piano mediano per vari valori dell'angolo di elevazione ϕ ([41]).

Modelli posizionali 3D – Head related transfer function (HRTF)



Vedere:

- Ian Vince McLoughlin, “Speech and Audio Processing”- Cambridge University Press (2016)
 - Cap. 4
- Aurelio Uncini “Audio digitale”, McGraw-Hill, 2006
 - Cap. 3.1, 3.2, 3.3.1-3.3.5, 3.3.7, 3.4, 3.5.1-3.5.2