unito.it
L'UNIVERSITÀ DI TORINO ON LINE

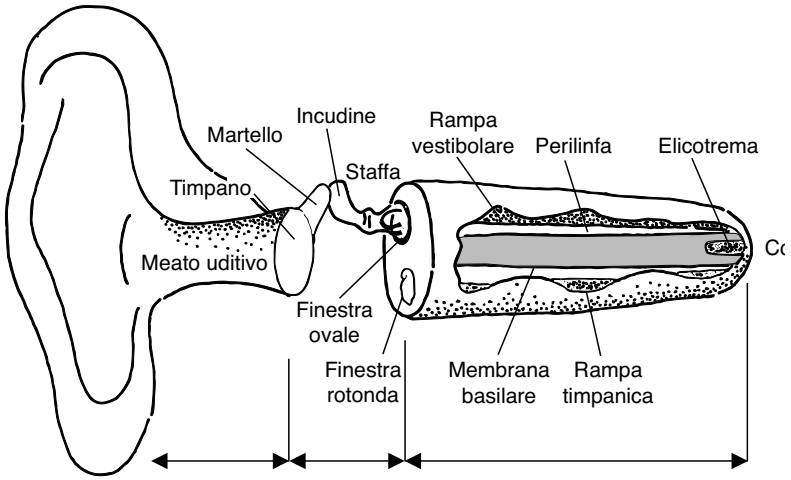
Tecnologie digitali per il suono e l'immagine 2020/21

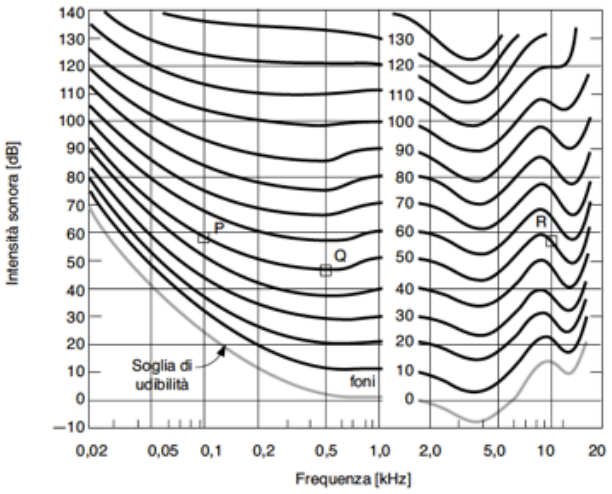
Vincenzo Lombardo
Corso di Laurea in DAMS
Università di Torino

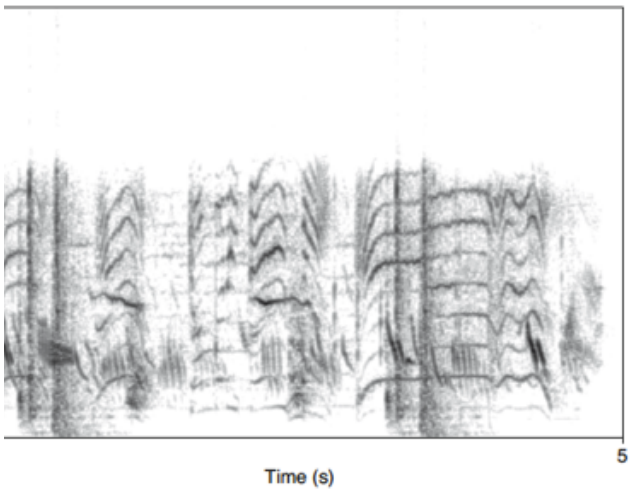
Mutuato in parte da Elaborazione audio e musica
(Laurea Magistrale di Informatica)

unito.it
L'UNIVERSITÀ DI TORINO ON LINE

La percezione uditiva







La natura dei suoni

- suoni come vibrazioni in un mezzo
- oscillazioni di pressione che variano per ampiezza e frequenza
- qualsiasi suono, discorso o sinfonia, è un'unica onda di pressione dell'aria

"s"

Fisica e cognizione

- In quanti modi si può descrivere un suono ?
  - Forte, debole, fragoroso, flebile, ... (volume)
  - Alto, basso, acuto, grave, baritonale ... (altezza)
  - Vuoto, pieno, brillante, opaco, metallico, plastico, ... (timbro)

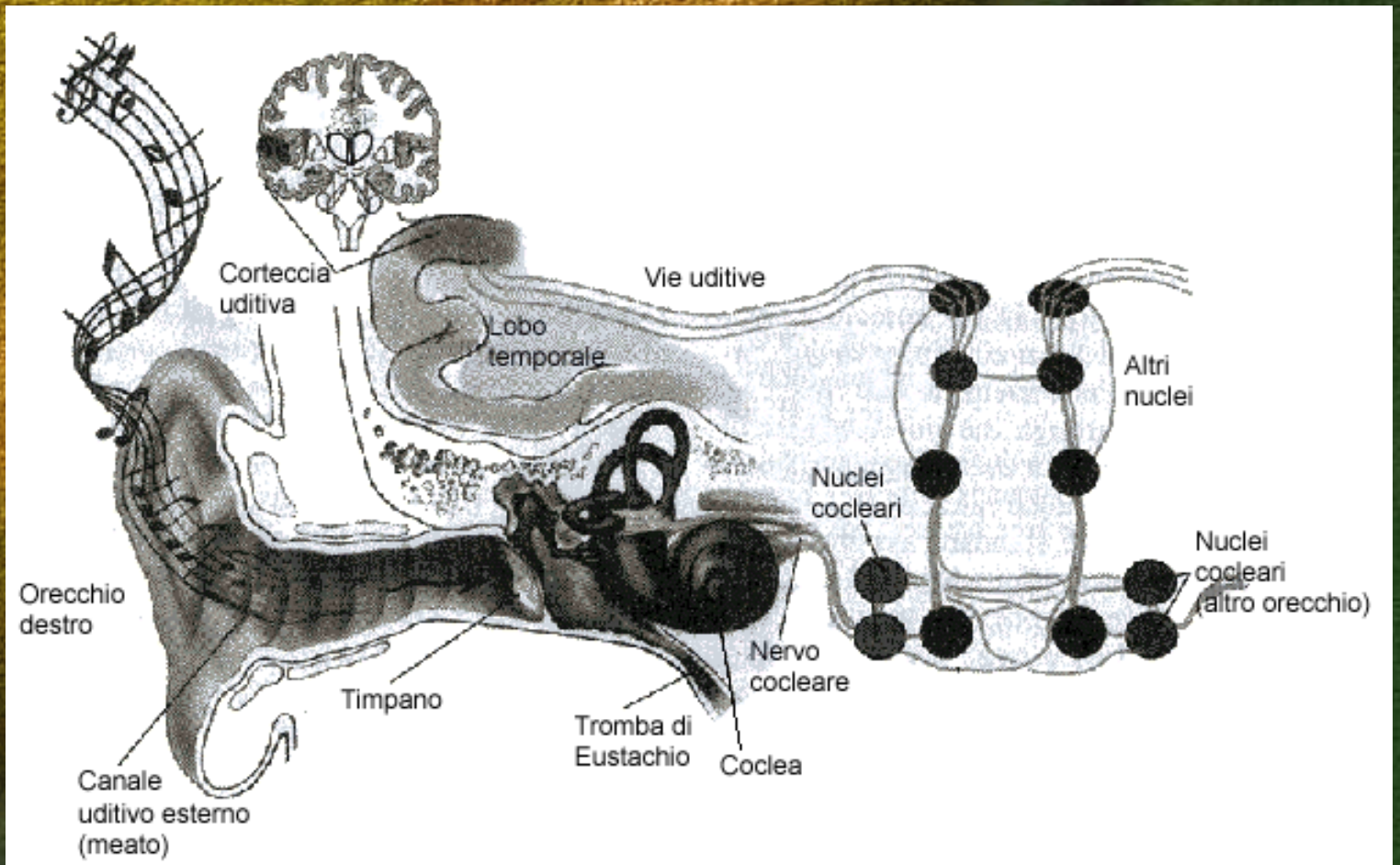
- Come fa l'uomo a percepire?

La percezione uditiva

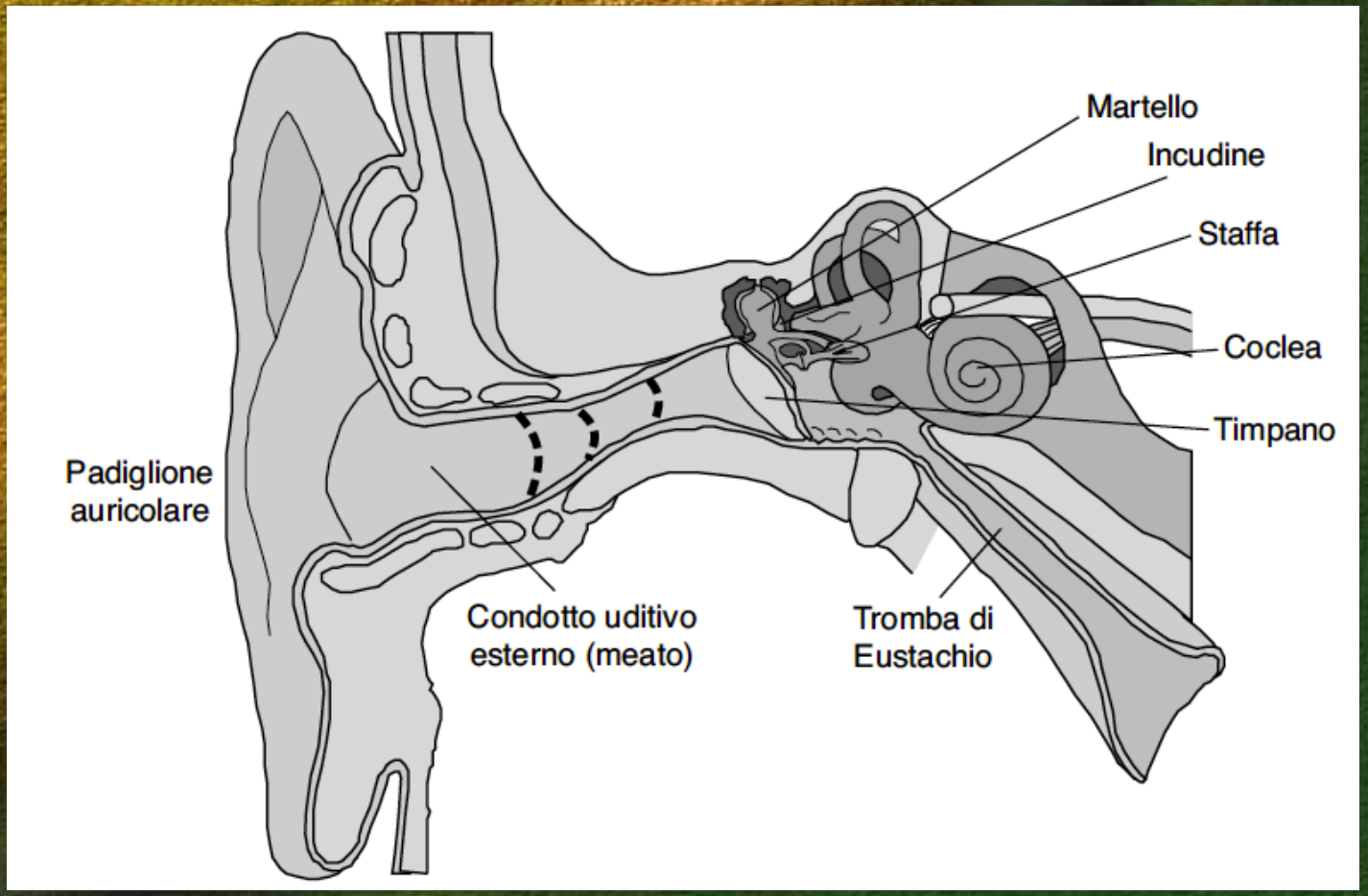
- Stimolo: energia acustica (onde sonore)
- Funzioni principali
  - comunicazione uditiva (tra cui il linguaggio)
  - localizzazione dei suoni (spazializzazione)

- Percezione uditiva = fisiologia dell'orecchio interno + azione del cervello

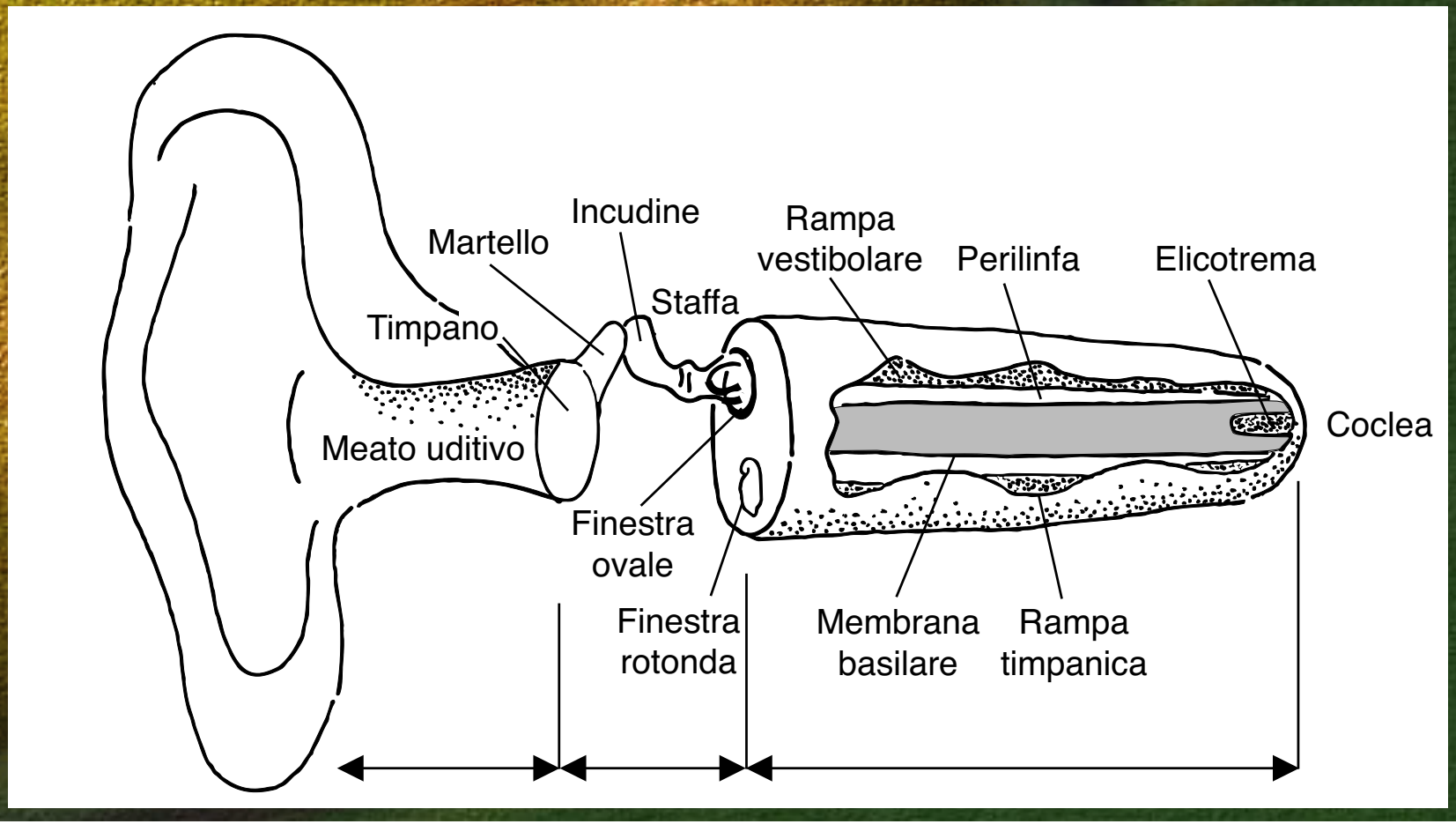
Fisiologia dell'udito



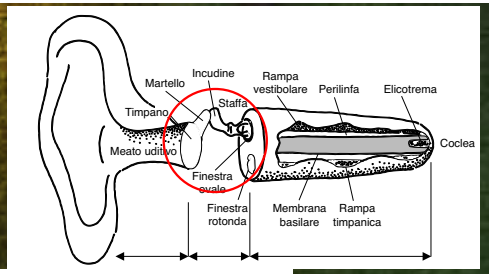
L’orecchio umano

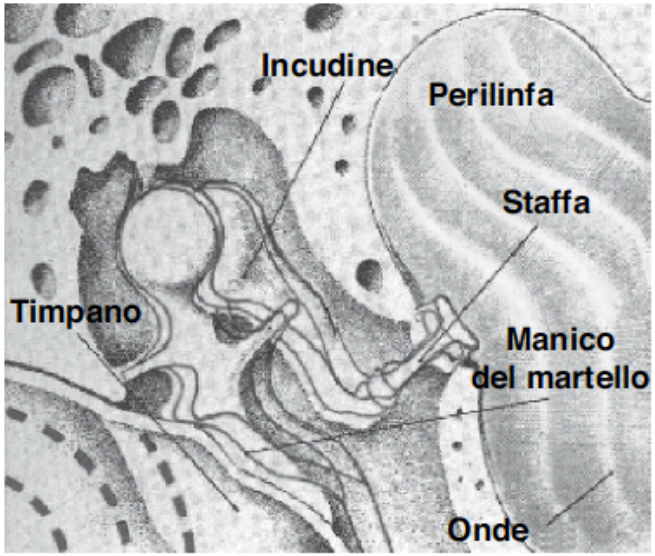


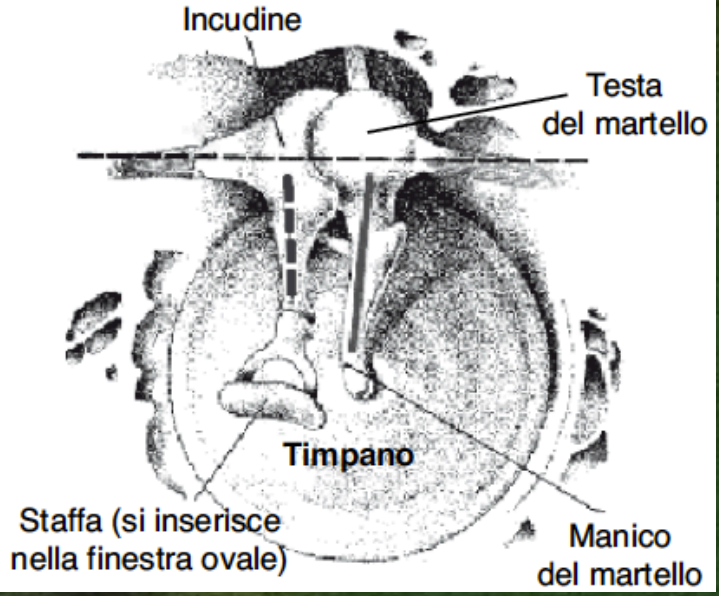
Versione schematica dell'orecchio



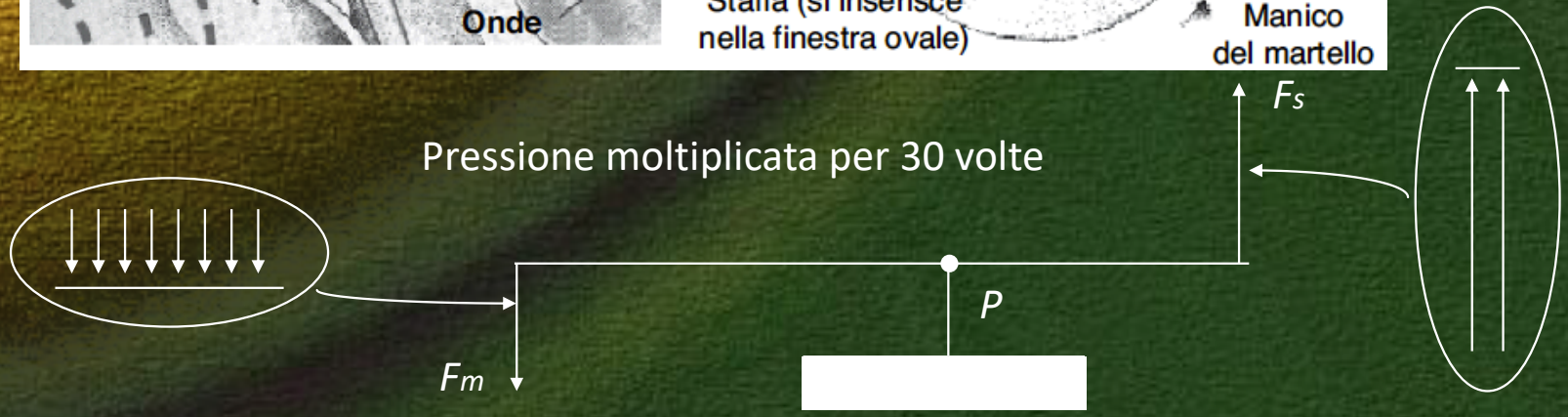
Gli ossicini e la leva



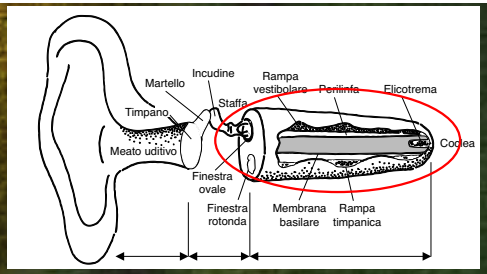


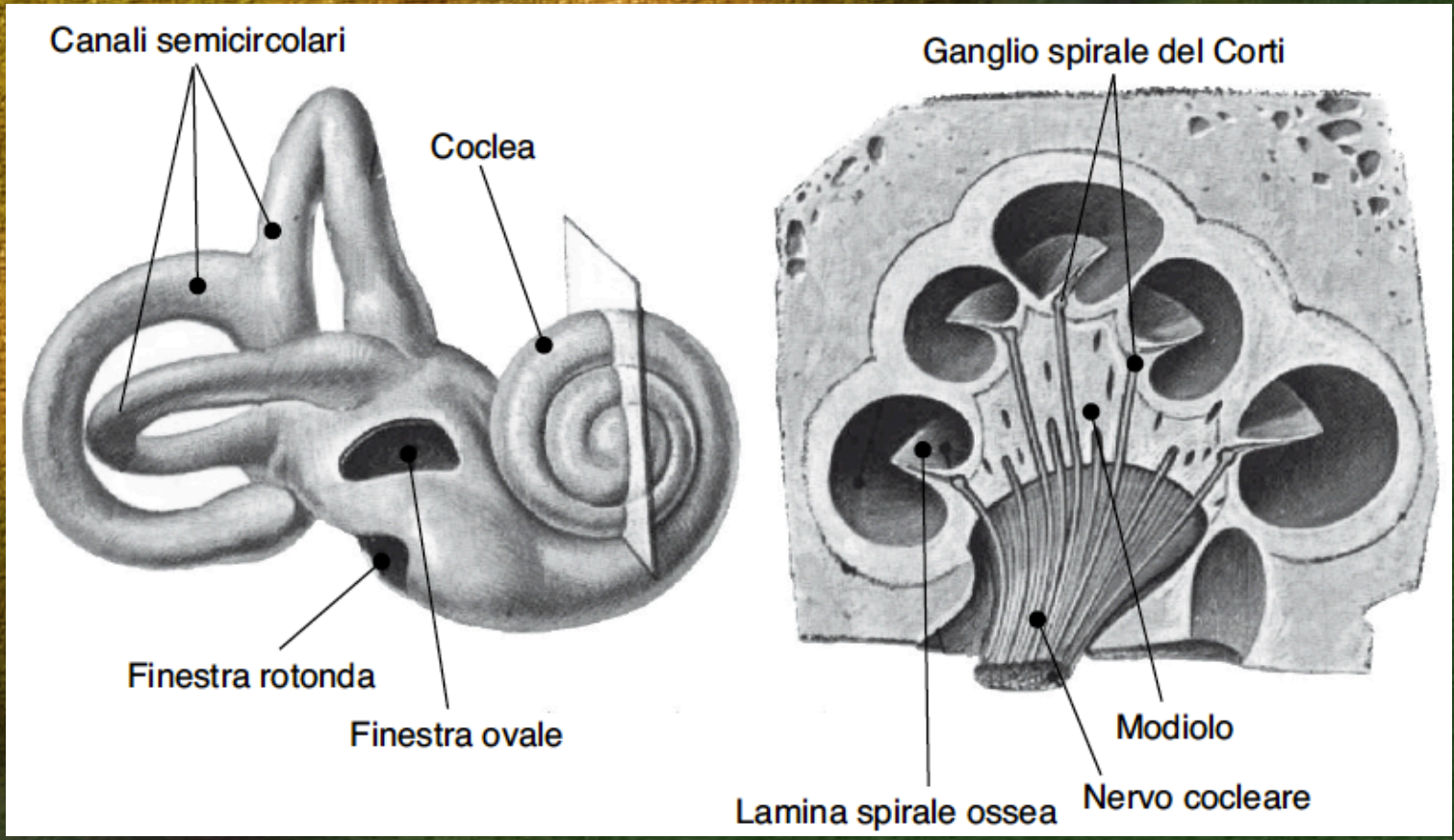


Pressione moltiplicata per 30 volte

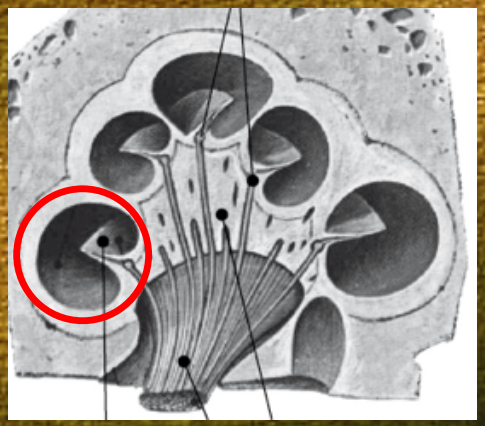


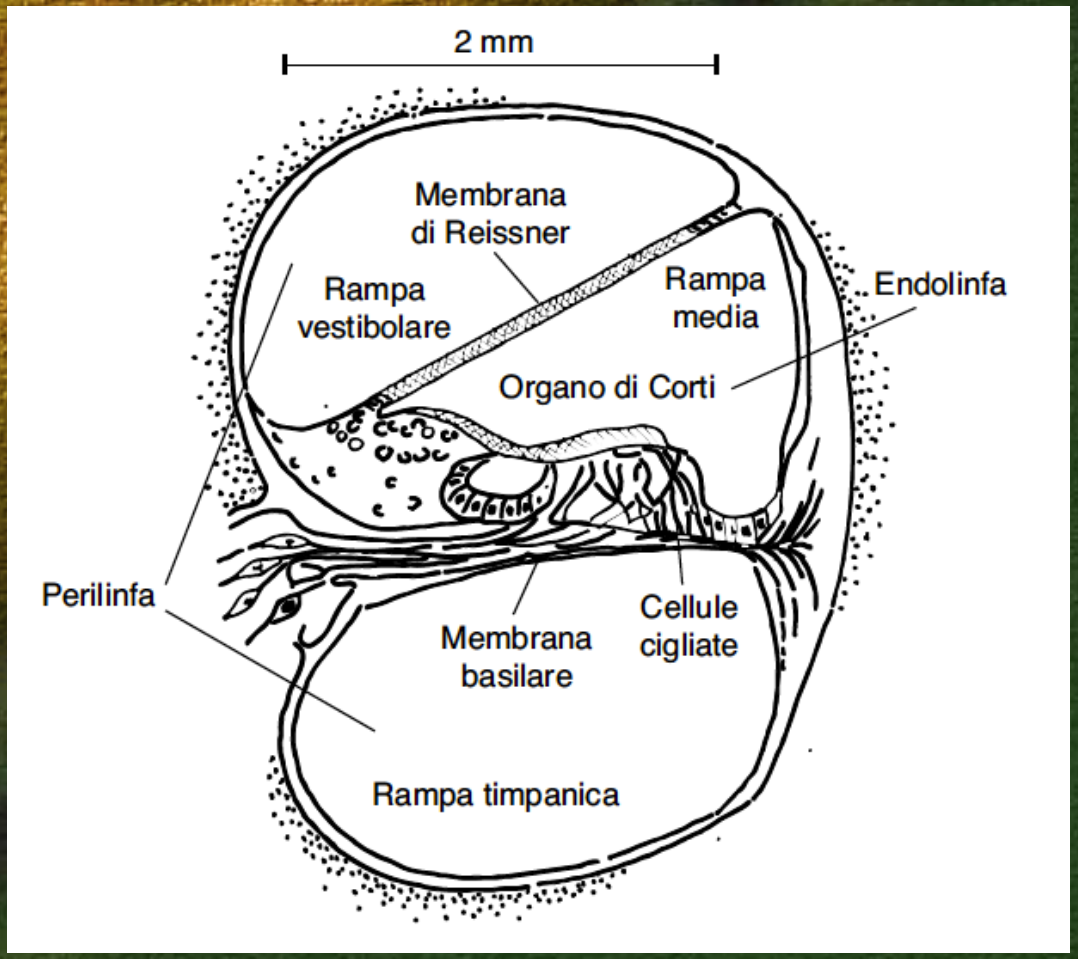
L'orecchio interno: la coclea



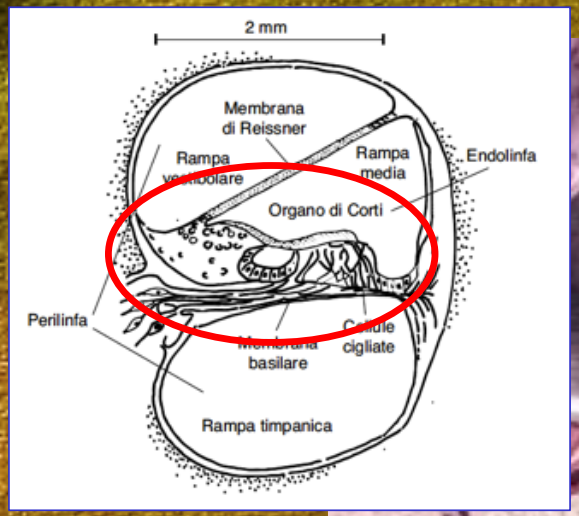


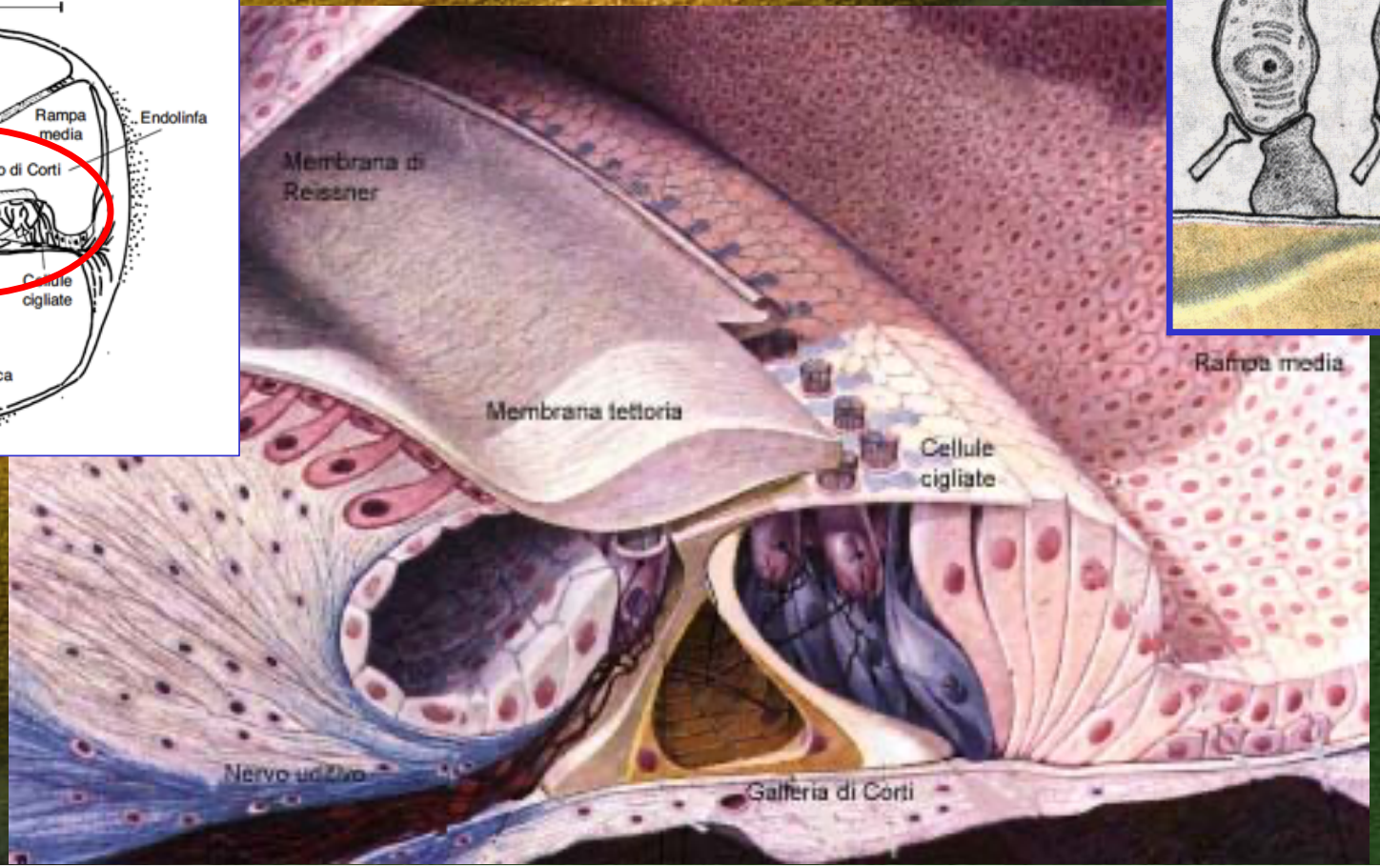
L’orecchio interno: la coclea

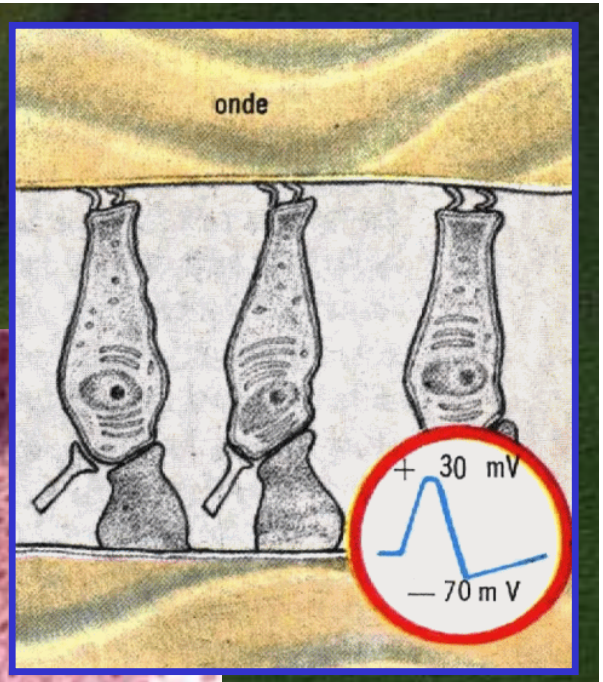




L’organo di Corti

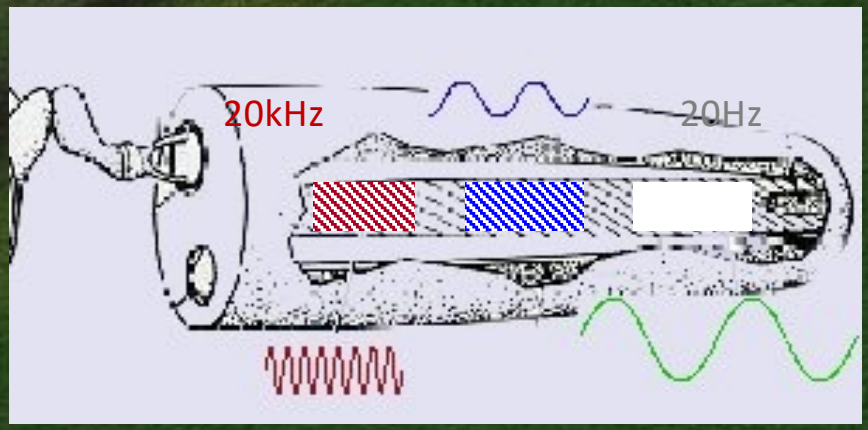




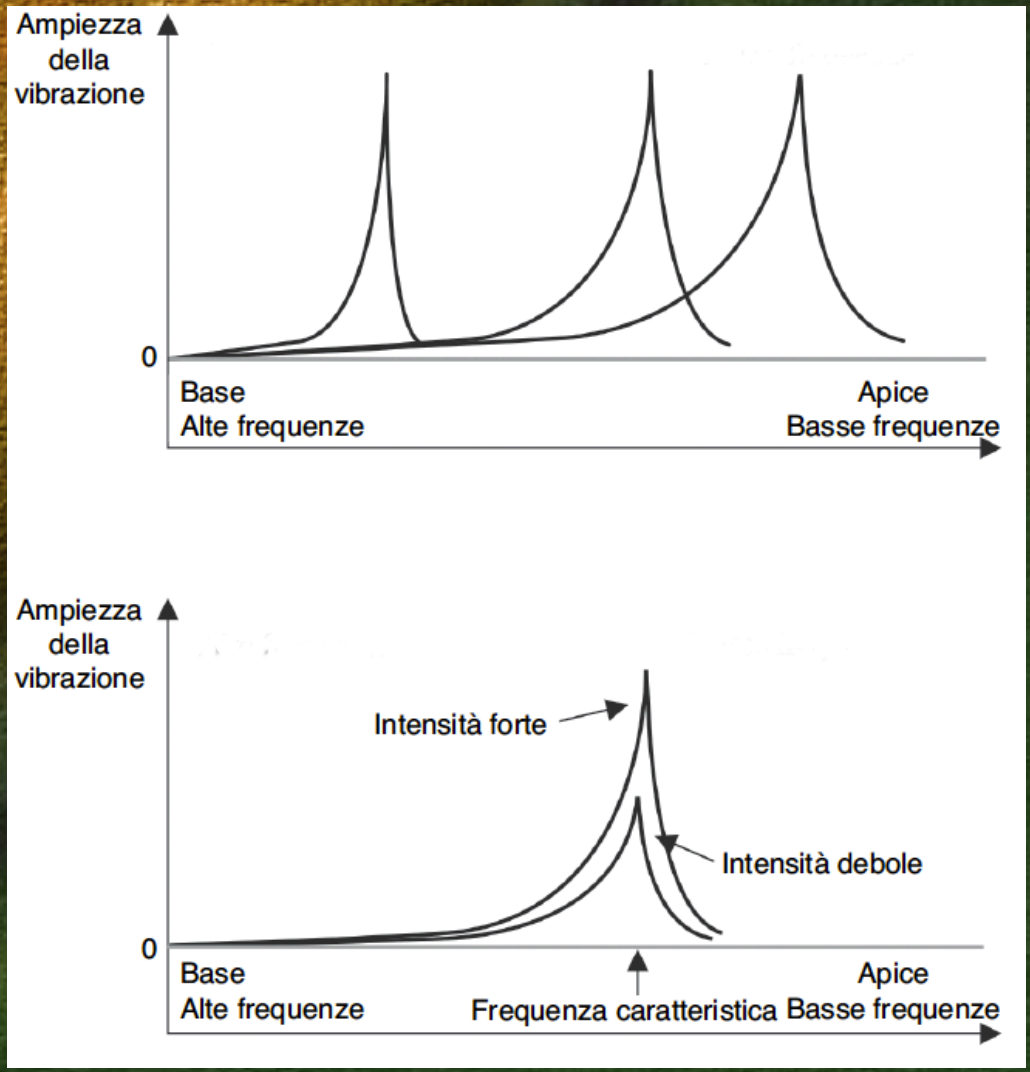


Membrana basilare

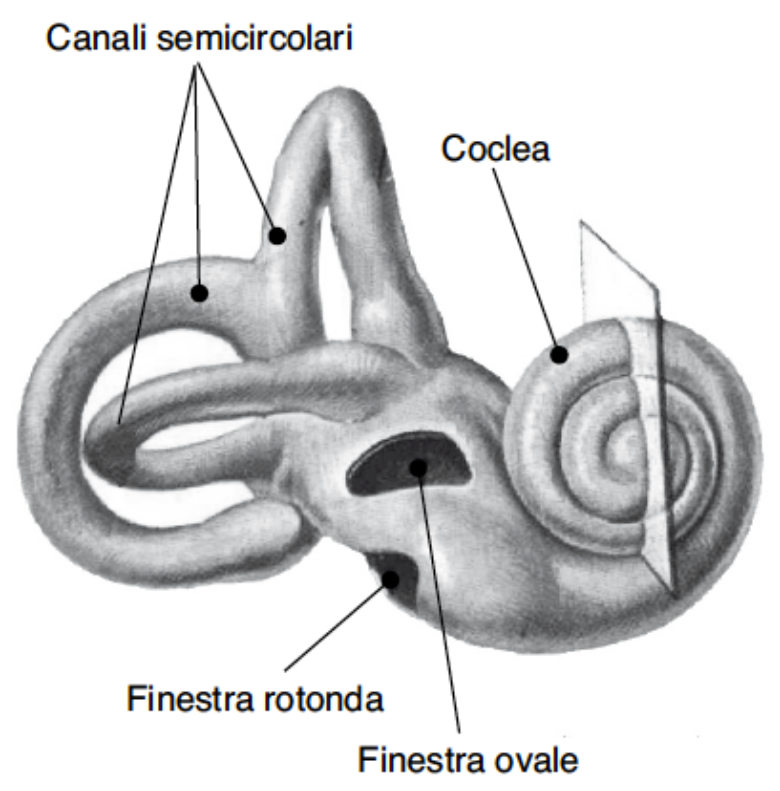
- Frequenze distribuite ordinatamente
  - alte all'estremità degli ossicini (stretta, rigida, leggera)
  - basse all'estremità interna (ampia, flessibile, massiccia)
- Banco di filtri accordati per bande
  - tutta la membrana copre l'estensione dell'udibile
  - un analizzatore di Fourier



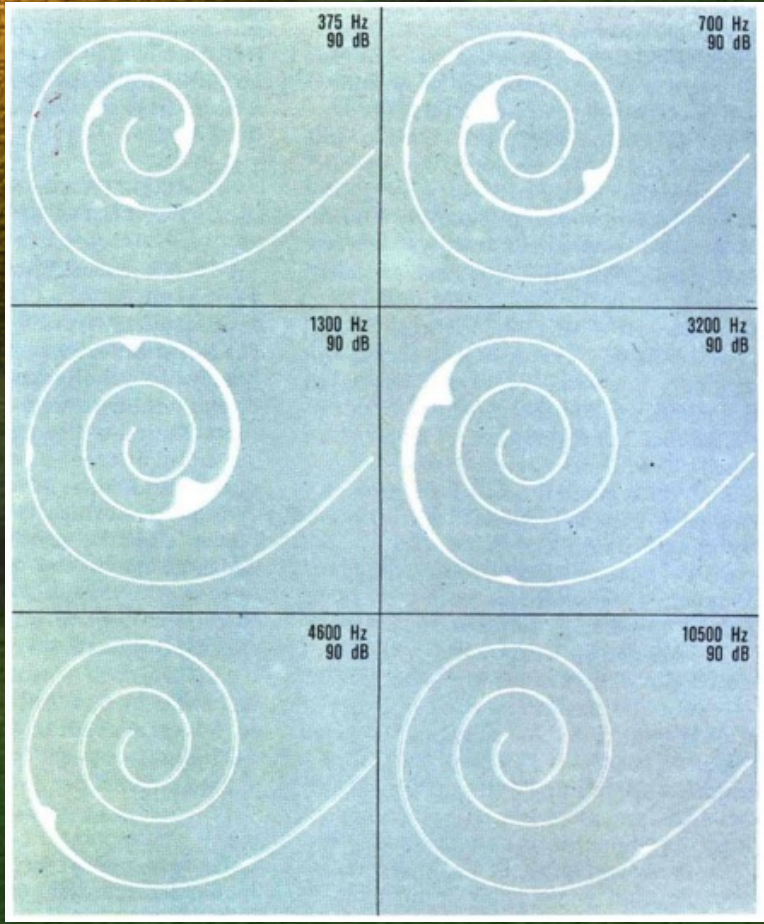
Vibrazione della membrana



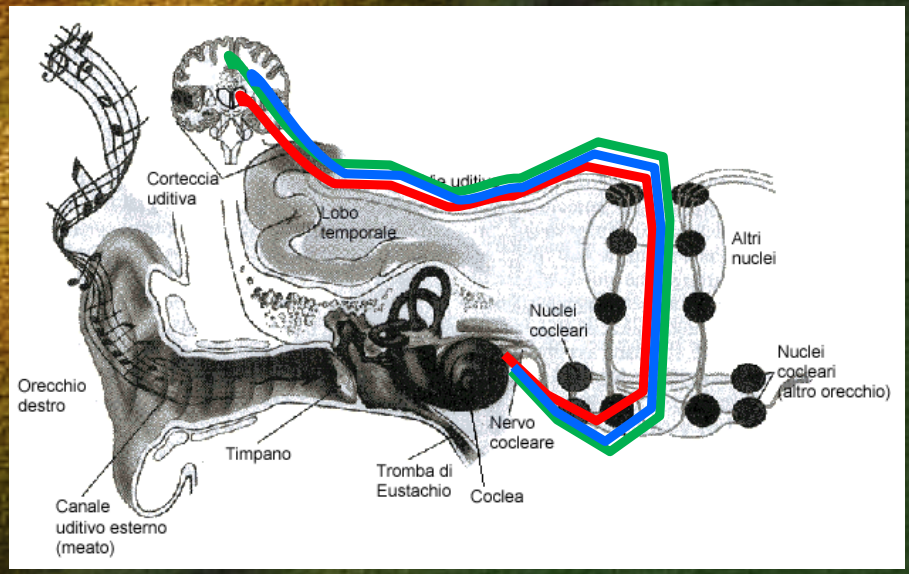
Funzionamento tonotopico



Studi elettro-fisiologici



Tassi di impulso



Il cervello rileva differenze tra

tasso spontaneo e tasso delle vibrazioni

- Tasso cresce con intensità della vibrazione
- Ciascuna fibra nervosa con frequenza caratteristica (stimolata con minima energia)

Azione di gruppo

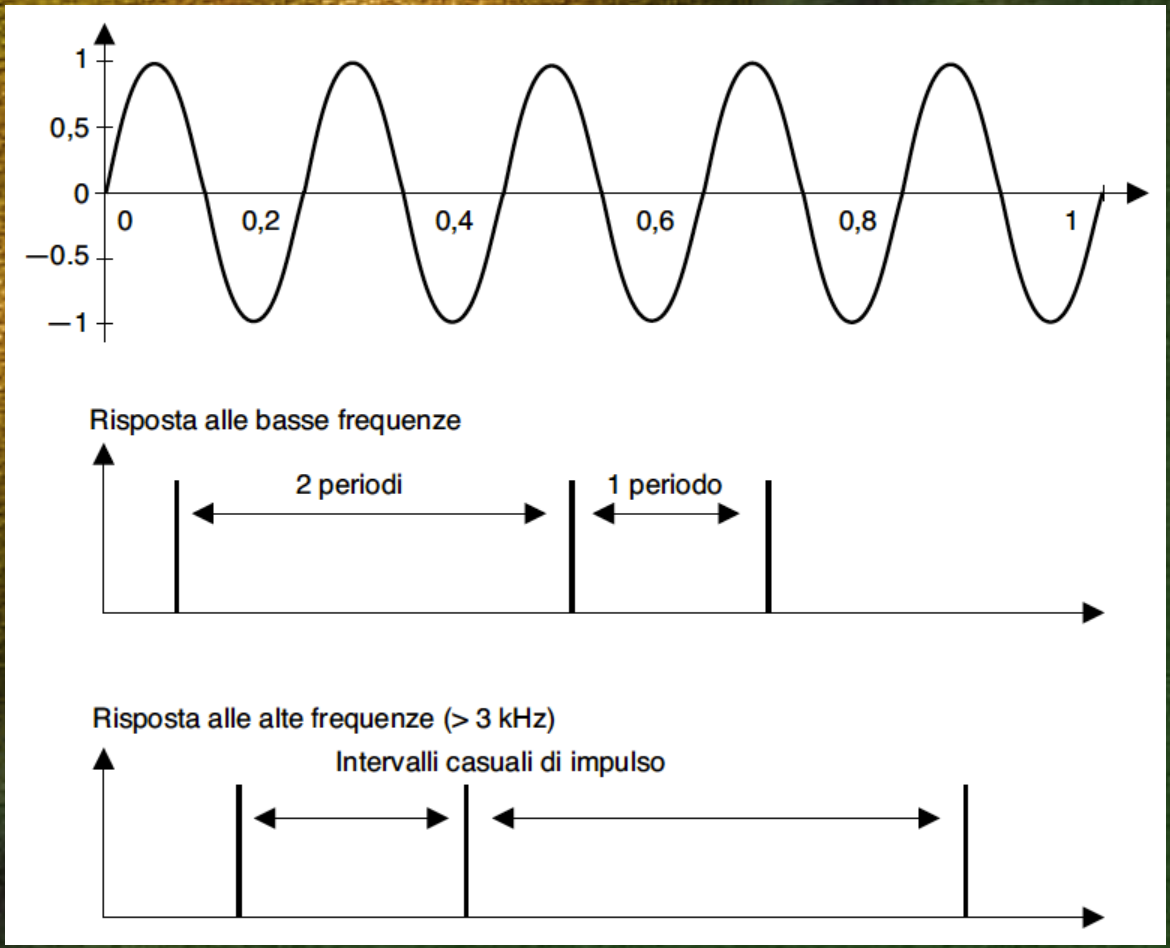
- Gran numero di cellule per una regione
- Cellule cercano di scaricare in sincrono con le vibrazioni (agli impulsi di picco)
  - Nota Do_2 di frequenza circa 131 Hz
  - Le cellule nervose scaricano 131 volte x sec
  - A frequenze superiori a 200 Hz non si può
- Metafora dei soldati in batterie

Phase locking, finché si può

Un segnale
in input

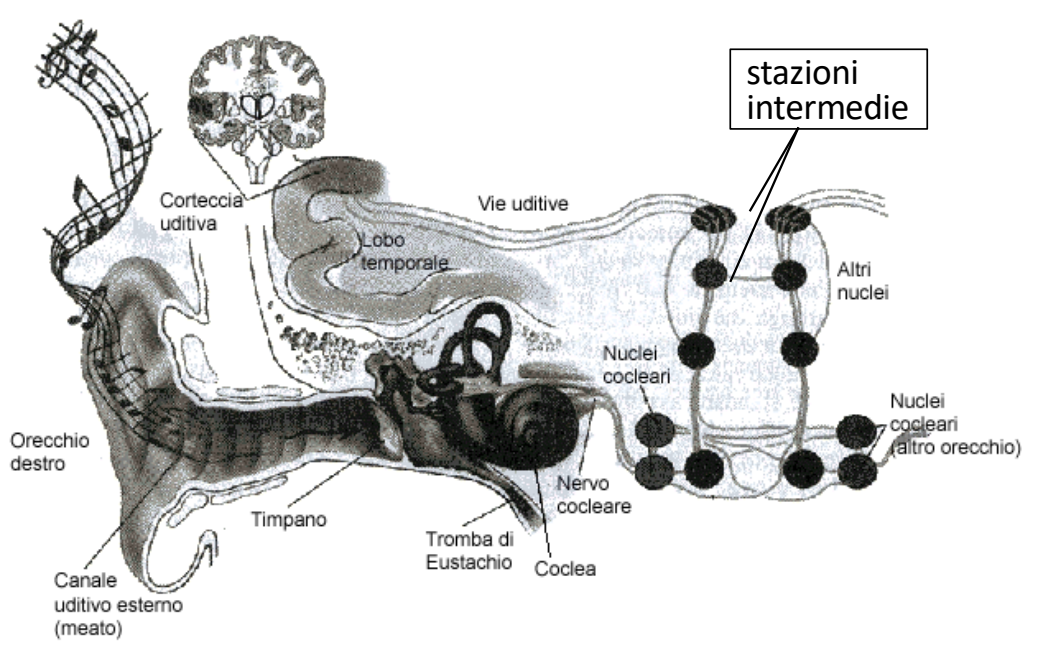
Anche
rilevatore di fase

Solo
discriminatore
di frequenze



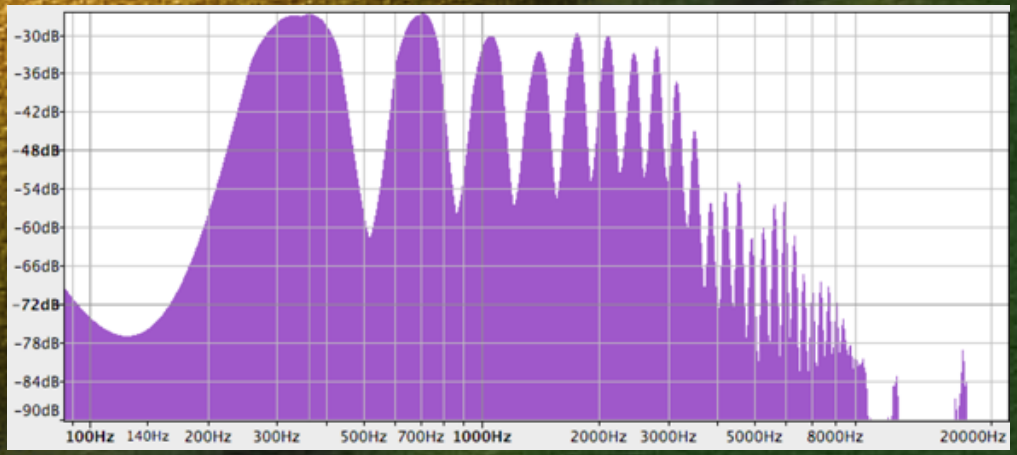
Dall’orecchio al cervello

- Segnali miscelati e elaborati in più stazioni
- Interpretazione nella corteccia uditiva



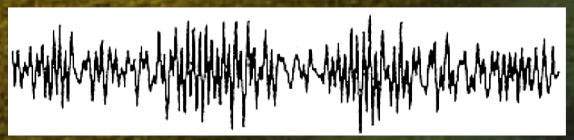
Rappresentazione duale del segnale

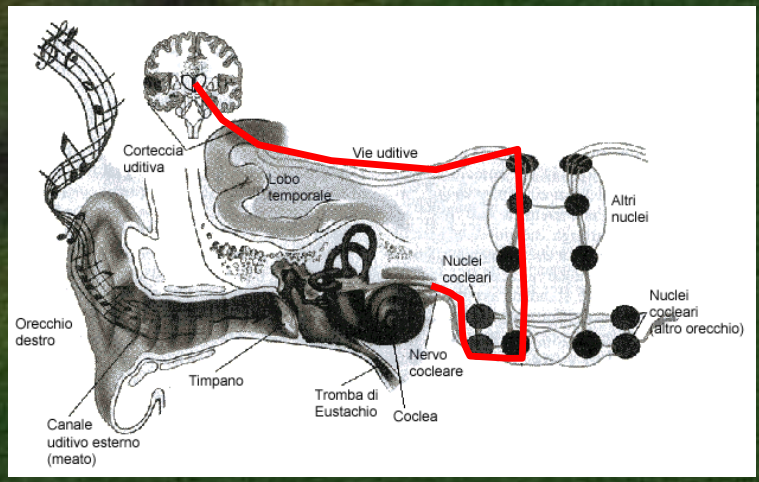
- Dominio della frequenza: organizzazione tonotopica



Rappresentazione duale del segnale

- Dominio del tempo
  - pattern di attivazione lungo il nervo uditivo
  - struttura della forma d'onda, transitori, inviluppo
  - sincronizzazione tra frequenze e tra stimoli dalle due orecchie





Conclusioni sulla fisiologia dell'udito

La nostra comprensione di come il cervello trasforma e interpreta l'informazione sul suono è ancora rudimentale


La psicologia dell’udito
onde sonore → percezione uditiva → cognizione

Fisica-percezione-cognizione

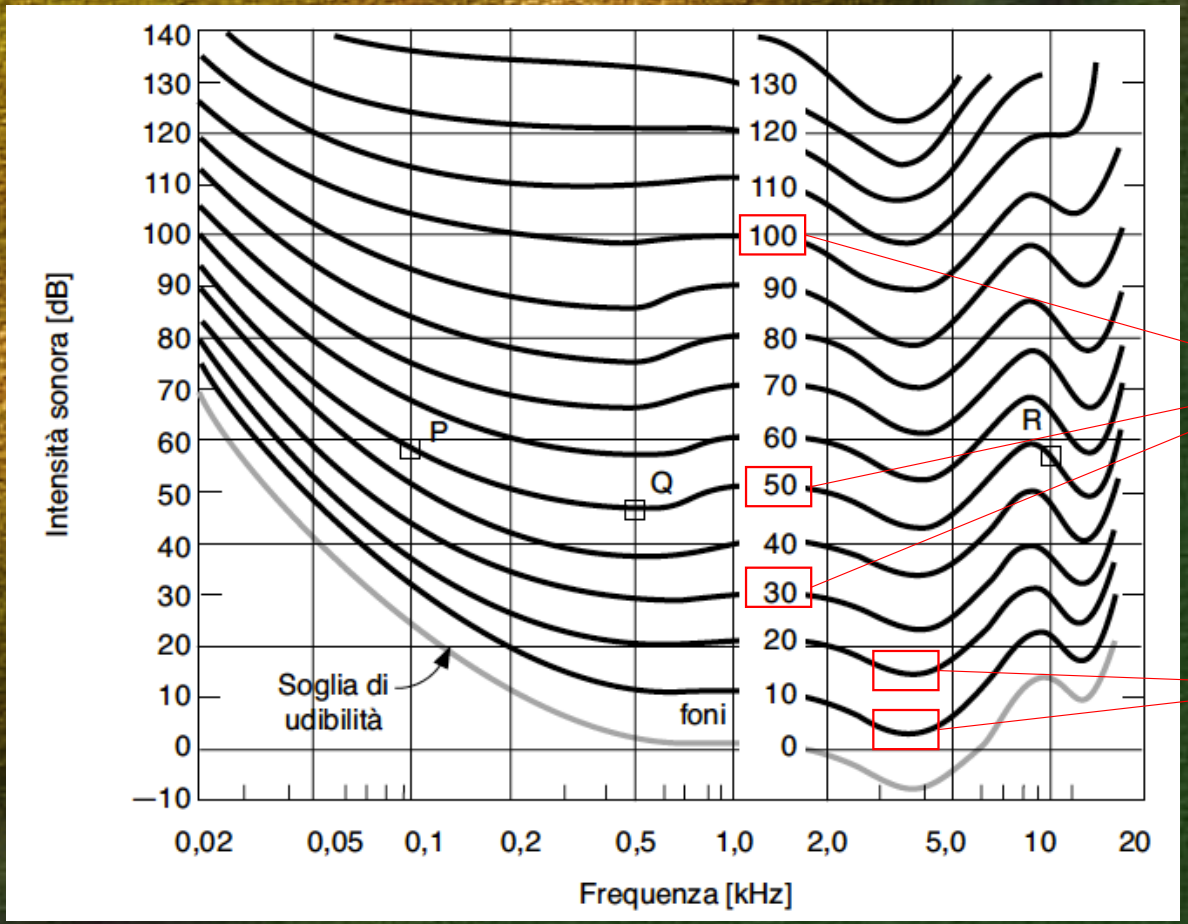
Aria	***Orecchio***	***Mente***
Suono	Sensazione uditiva	Musica/Parole
Ampiezza	Intensità	Dinamica
Frequenza	Altezza	Classe di toni
Spettro	Timbro	Riconoscimento sorgente
Propagazione	Localizzazione	Mappa spaziale soggettiva

Volume

- SIL in dB (10 log I / I_0)
  - $I_0 = 10^{-12}$ W/m^2
  - I_0 a 1000 Hz (soglia udibile per ascoltatori acuti)
- Volume percepito (LL) in foni (*phons*)
  - intensità che dipende dalla frequenza
  - dato un suono A, quanto è forte un suono B a 1000Hz che è forte uguale
- Volume soggettivo (L) in soni (*sones*)
  - qual è la differenza di L tra due suoni
  - 100 soni è percepito come il doppio di 50 soni

Diagramma di Fletcher-Munson

in pratica livelli
da 10 a 20 dB
(e superiori)
per frequenze
non centrali



valori di volume
a 1000 Hz

sensibilità max
per suoni
da 2 a 5 kHz

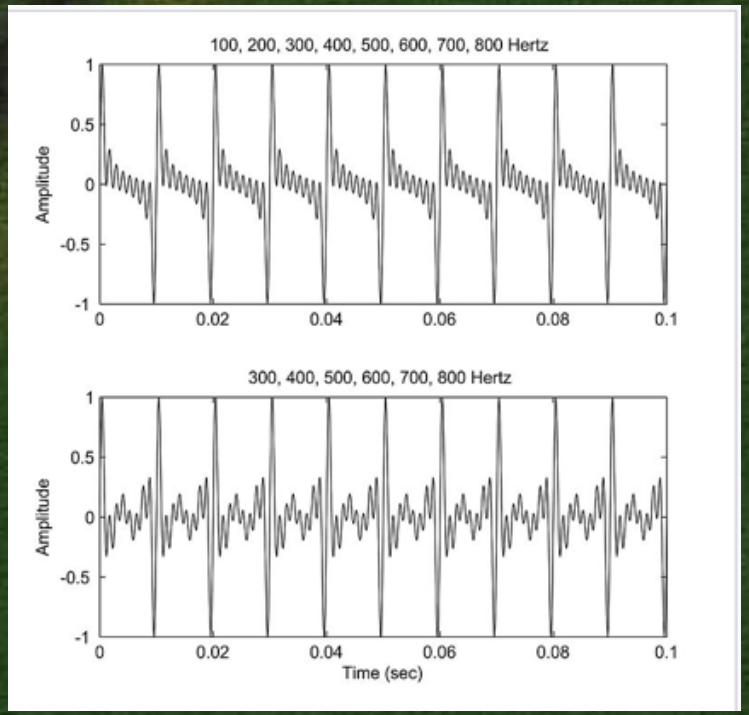
Frequenza e altezza

- tendenzialmente da 20 a 20000 Hz (20 kHz)
  - di solito fino a 17-18 kHz per un adulto in buona salute
  - vecchiaia: 12 Khz (donne), 5 Khz (uomini)
- Simile al problema ampiezza-volume
- Approssimazione: ottava come i bel (log in base 2 invece che 10)

Estremità e valori abituali

- suoni sotto 30 Hz piuttosto difficili da udire
  - forte intensità e isolamento per onde sin di 15 Hz
  - sotto i 20 Hz si passa al "sentire" (sopra 100 dB)
- in natura, non onde pure sinusoidali
  - Udibile anche Do_0 (16 Hz), organo a canne
  - Contributo delle armoniche
- altezza residua o frequenza fantasma

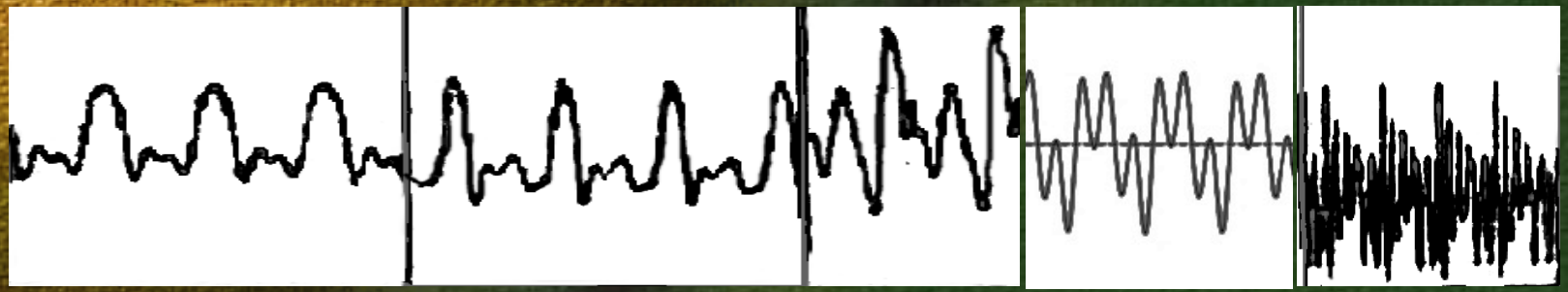
[Wikipedia: Missing Fundamental]



Timbro

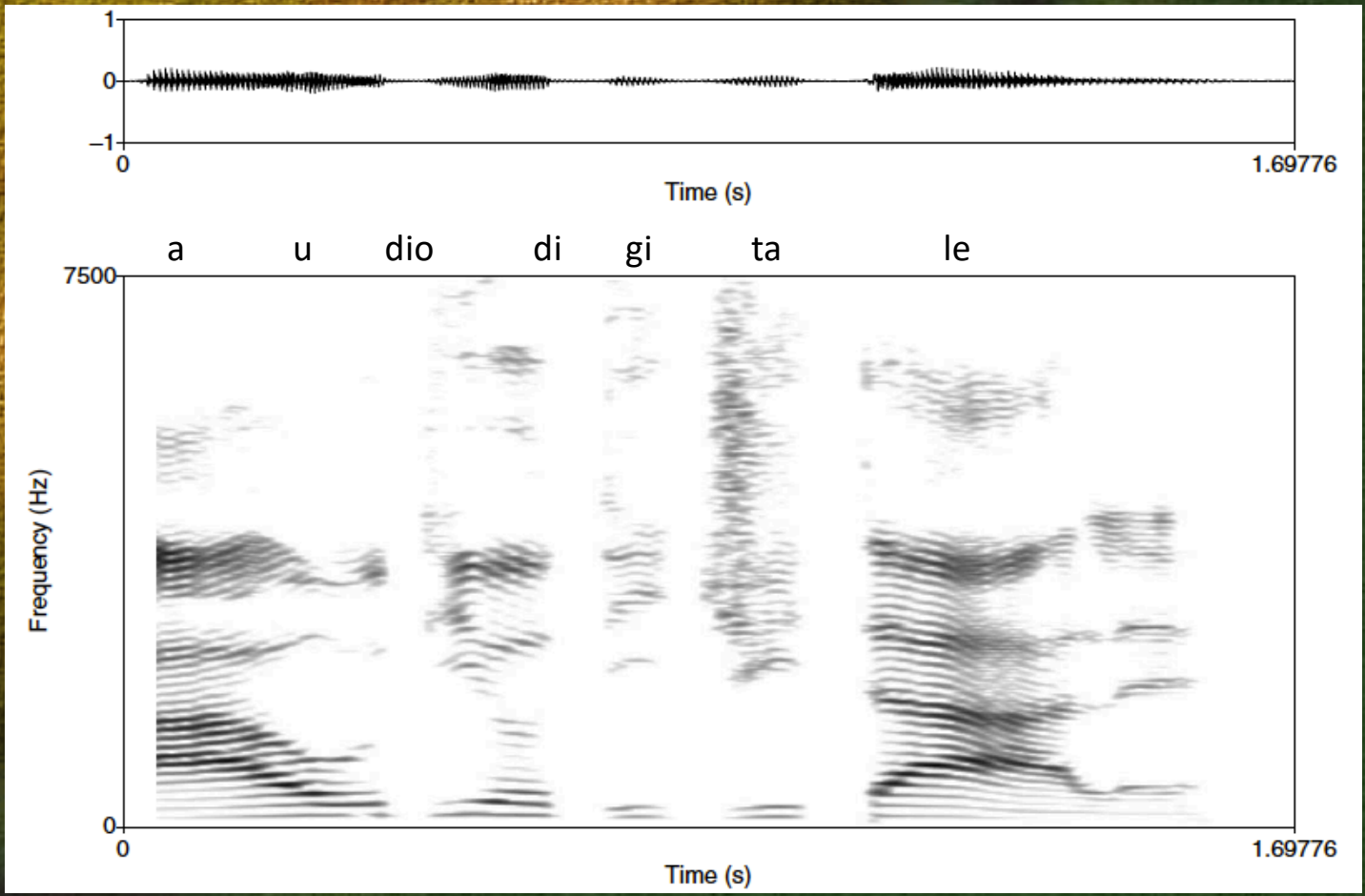
- Timbro dalla forma d'onda
- Relazione più difficile da trovare

[immagine adattata da Pierce]



flauto tromba sax soprano violino chitarra tuba

Formanti


1
0
−1
0
1.69776
Time (s)
a u dio di gi ta le
7500
Frequency (Hz)
0
0
1.69776
Time (s)

Timbro e livello dinamico: transitori

- Timbro cambia con il livello dinamico
- Esempio: tromba
  - più brillante il suono a forti intensità
  - forte tromba lontana VS debole tromba vicina
- Attenzione nella sintesi

Riconoscimento strumenti: transitori (attack e decay)

- Durata dei transitori varia tantissimo: dipende da strumento e esecutore
  - 20 ms per un oboe
  - 30-40 ms per tromba o clarinetto
  - 70-90 ms per flauto o violino

- Note sopra il Do centrale → periodo di 2-4 ms: il transitorio comprende più cicli di vibrazione

Riconoscimento strumenti: vibrato caratterizzante

- Vibrato = modulazione periodica dell’altezza di un suono
- Corrispondente in ampiezza: tremolo
- Suono sintetizzato: spesso assenza di vibrato realistico

Nyquist command per vibrato

http://forum.audacityteam.org/viewtopic.php?f=39&t=55334
per suoni di 2 sec)

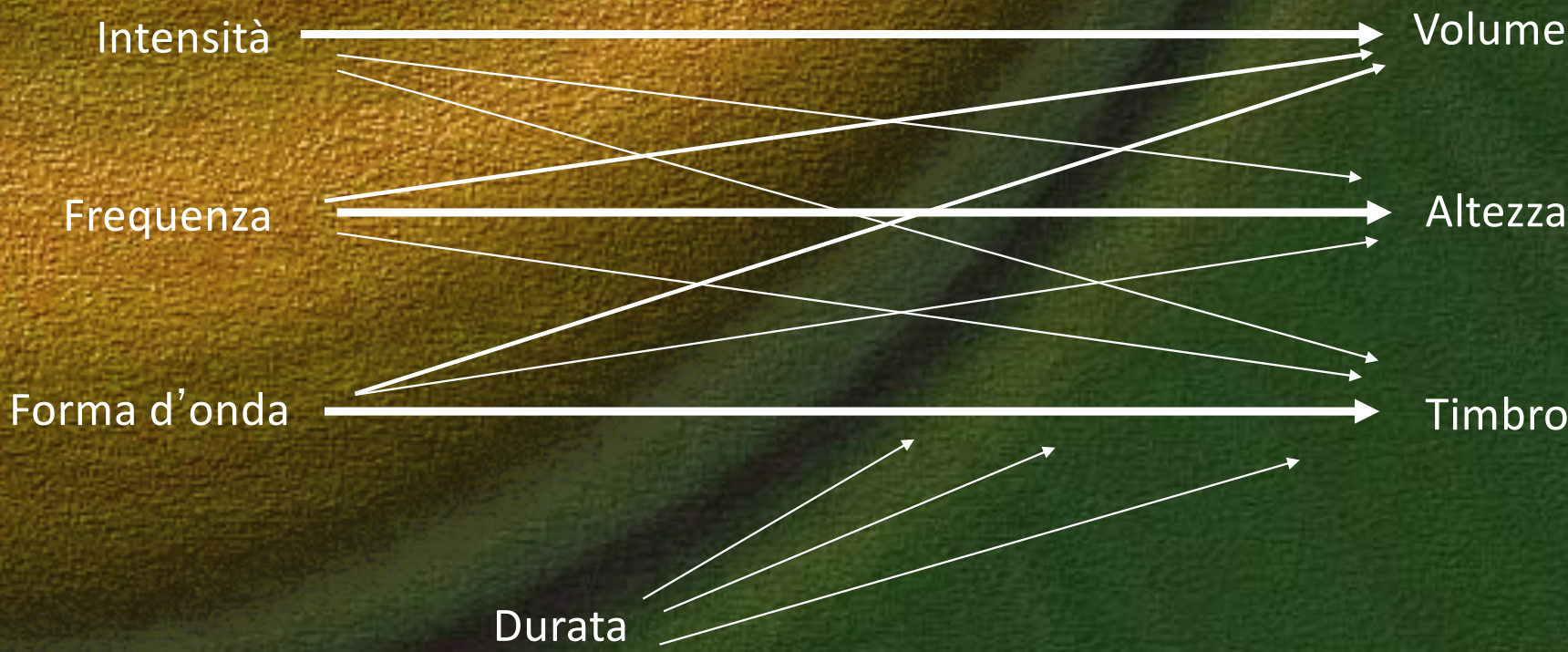
Copiare nel prompt Nyquist di Audacity (menù Strumenti)

```
(setq initial-speed 1.0) ; Hz
(setq final-speed 4.0)
(setq initial-depth 50) ; scale of 0 to 100
(setq final-depth 50)

(setq initial-speed (/ initial-speed 2.0)) ; 1/2 the vibrato speed
(setq final-speed (/ final-speed 2.0))
(setf vib-speed (pwlv initial-speed 1 final-speed)) ; the vibrato speed envelope

(setq initial-depth (/ initial-depth 100.0));changes from % to a scale of 0 to 1
(setq final-depth (/ final-depth 100.0))
(setf vib-depth (pwlv initial-depth 1 final-depth)) ;the vibrato depth envelope

(setf *s-table* (list s 0 t)) ; makes the sound 's' into a wavetable

(fmosc 0.0 (mult vib-depth (fmosc 0 vib-speed)) *s-table* 0)
```

Riconoscimento strumenti: differenze di attacco

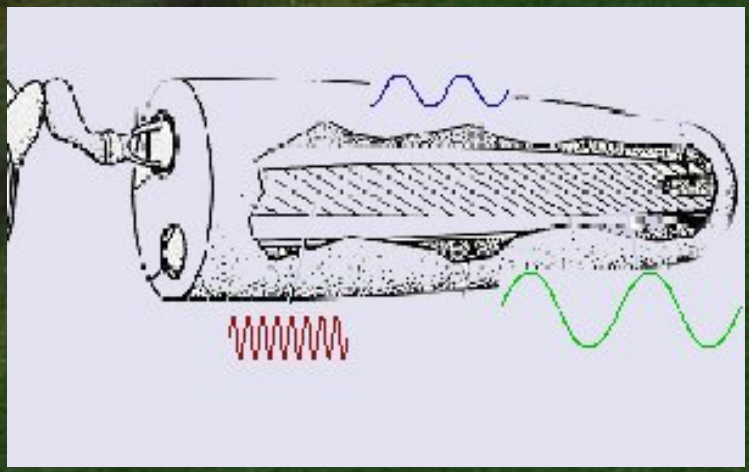
- Alta sensibilità alle differenze dei suoni tra le due orecchie
- Si percepiscono ritardi di pochi microsecondi tra due suoni

Rapporti fisica-percezione



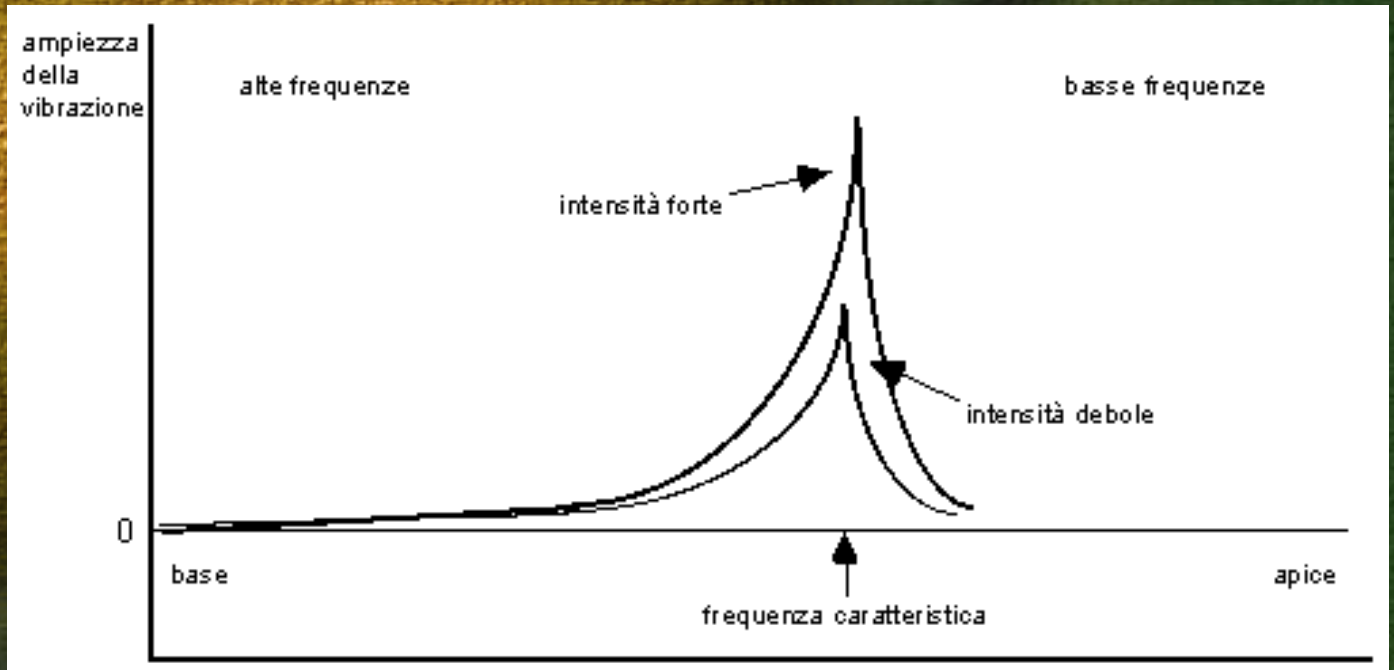
L’interferenza tra i suoni: mascheramento

- Funzionamento della membrana basilare
- Siamo in natura (non in matematica)
  - la regione del picco ha una dimensione
  - incertezza nella percezione dell’altezza di un suono



La causa del mascheramento

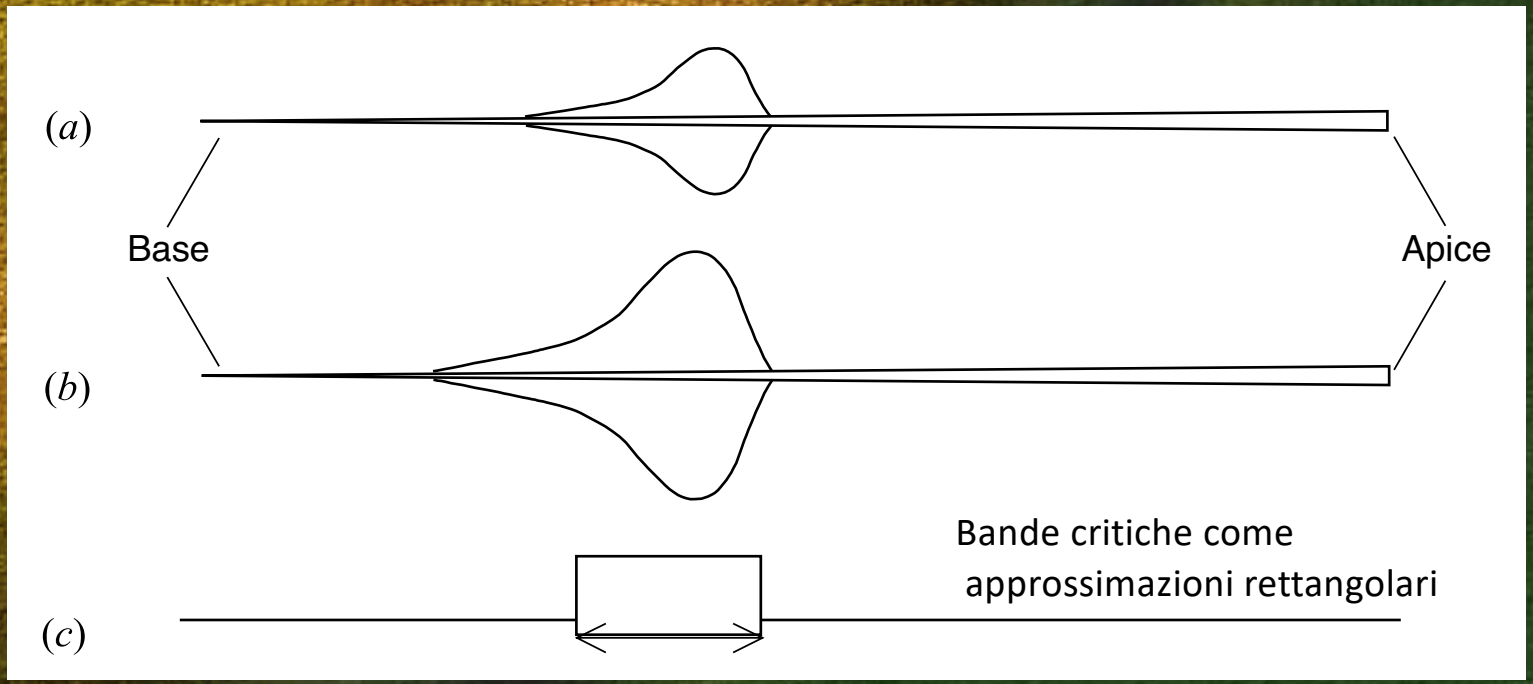
- I neuroni si “bloccano” per scaricare assieme al picco del segnale
- Coclea = phase-detector + frequency-discriminator



Nella vita quotidiana

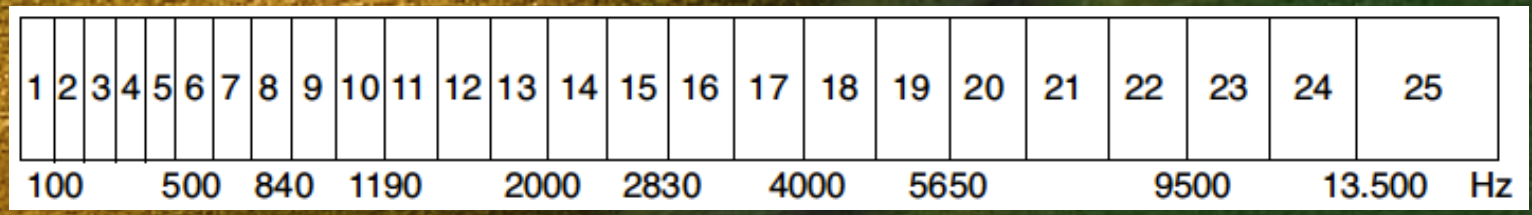
- Non si riesce ad ascoltare qualcuno che bisbiglia dove qualcun altro sta urlando
- E' analogo all'effetto "cattura" nella radio FM

Bande critiche di Fletcher



Bande critiche di Fletcher

Bande critiche come
approssimazioni rettangolari

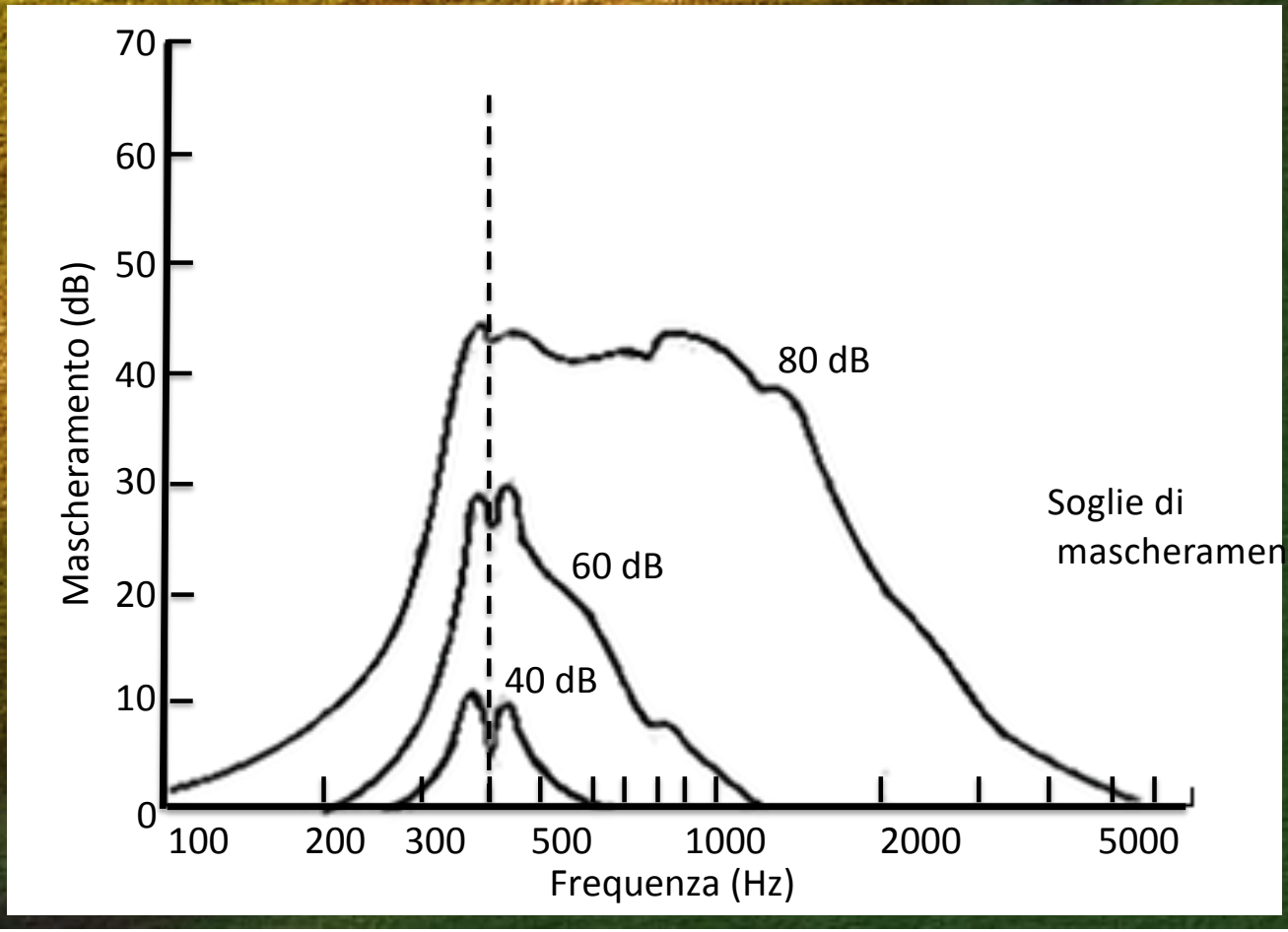


Bande critiche di Fletcher

Numero di banda	Centro della banda	Estremi della banda e estensione
1	60	fino a 100 (80)
2	150	100-200 (100)
3	250	200-300 (100)
4	350	300-400 (100)
5	450	400-500 (100)
6	550	500-600 (100)
7	655	600-710 (110)
8	775	710-840 (130)
9	920	840-1000 (160)
10	1095	1000-1190 (190)
11	1300	1190-1410 (230)
12	1545	1410-1680 (270)
13	1840	1680-2000 (320)
14	2190	2000-2380 (380)
15	2605	2380-2830 (450)
16	3095	2830-3360 (530)
17	3680	3360-4000 (640)
18	4380	4000-4760 (760)
19	5205	4760-5650 (890)
20	6175	5650-6720 (1.050)
21	7360	6720-8000 (1.280)
22	8750	8000-9500 (1.500)
23	10.400	9500-11.300 (1.800)
24	12.400	11.300-13.500 (2.200)
25	16.700	da 13.500 (6.500)

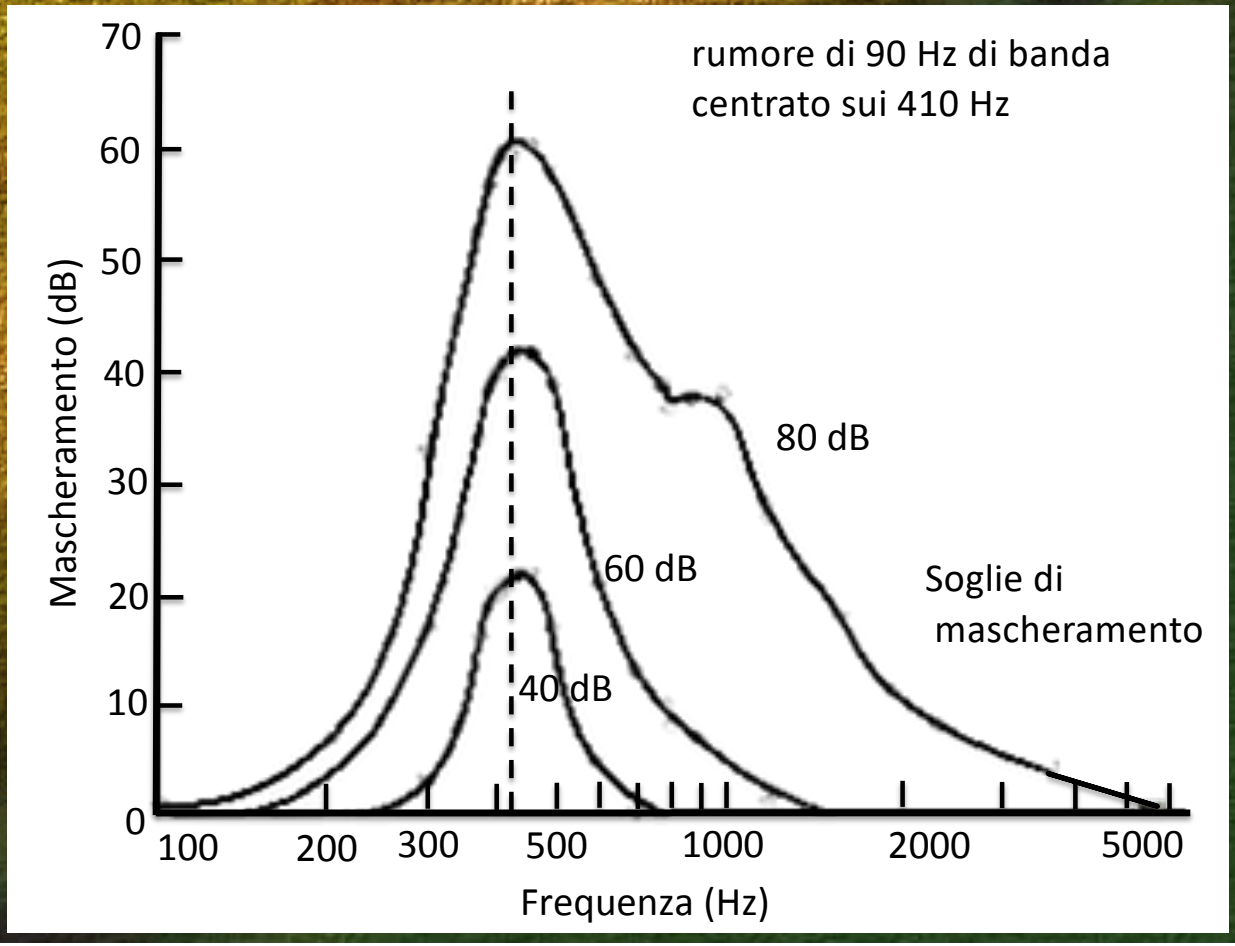
Il mascheramento tonale

[immagine adattata da Pierce]



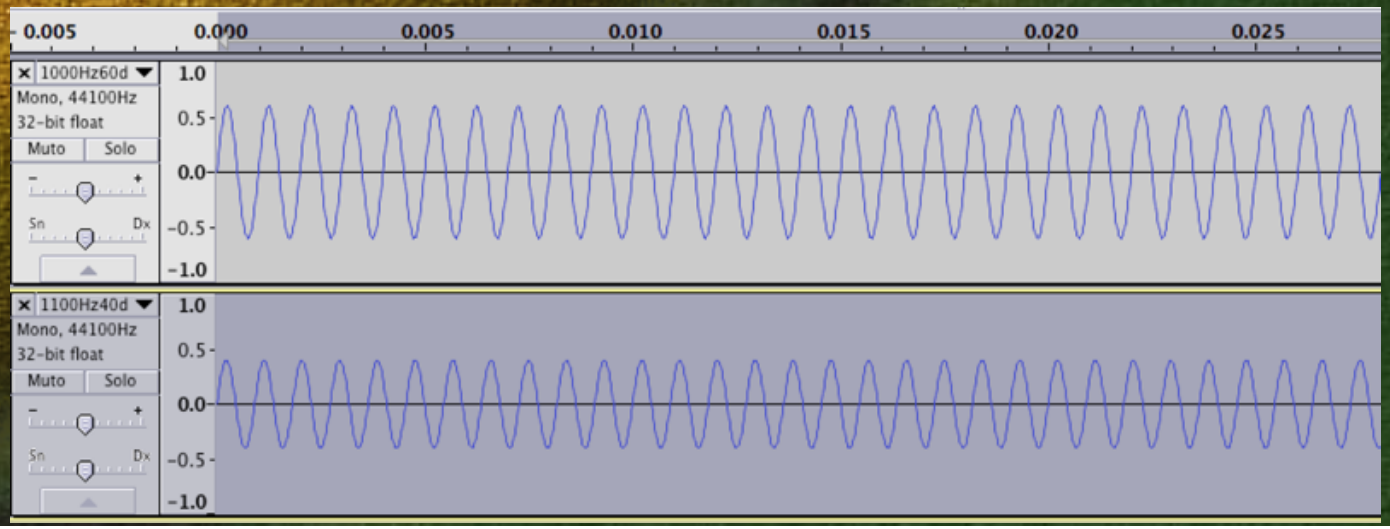
Il mascheramento non tonale

[immagine adattata da Pierce]



Mascheramento temporale

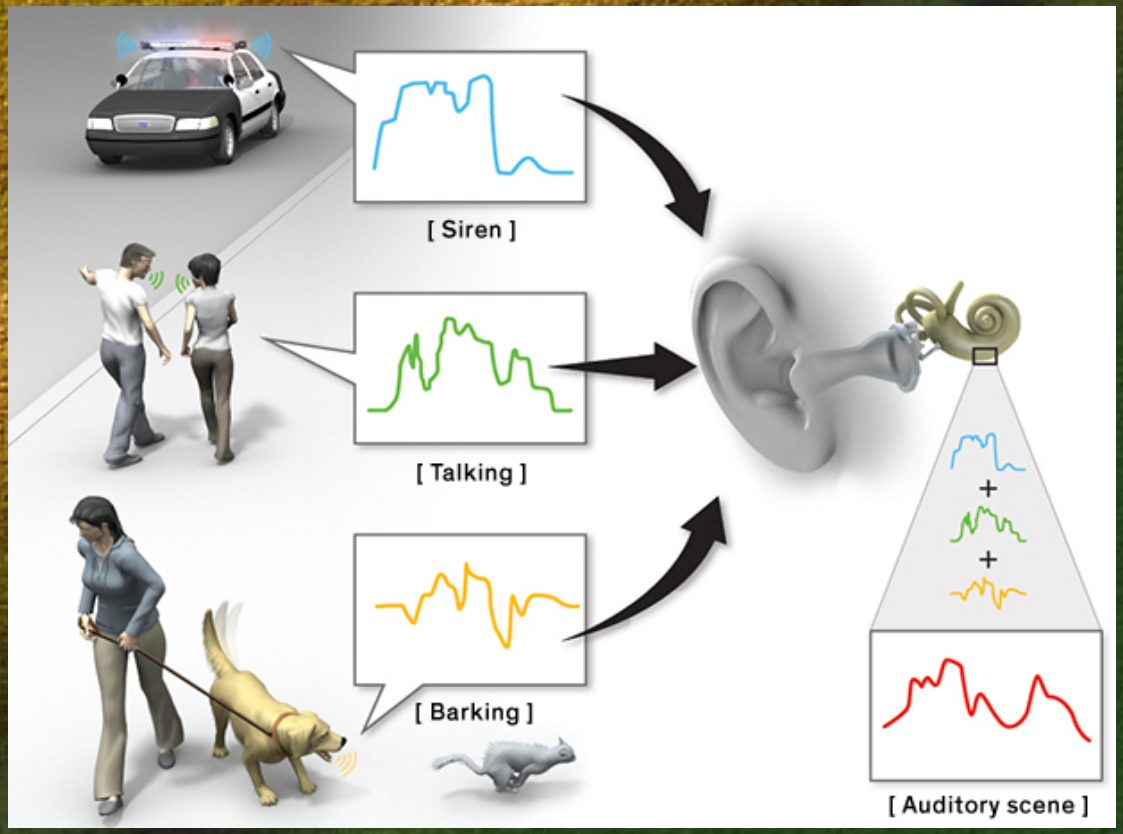
- mascheramento oltre la durata del mascheratore
- Esempio: A (tono 1000 Hz / 60 dB) maschera B (tono 1100 Hz / 40 dB) per 5 msec oltre A
- ritardo maggiore con B più debole



Organizzazione percettiva del suono

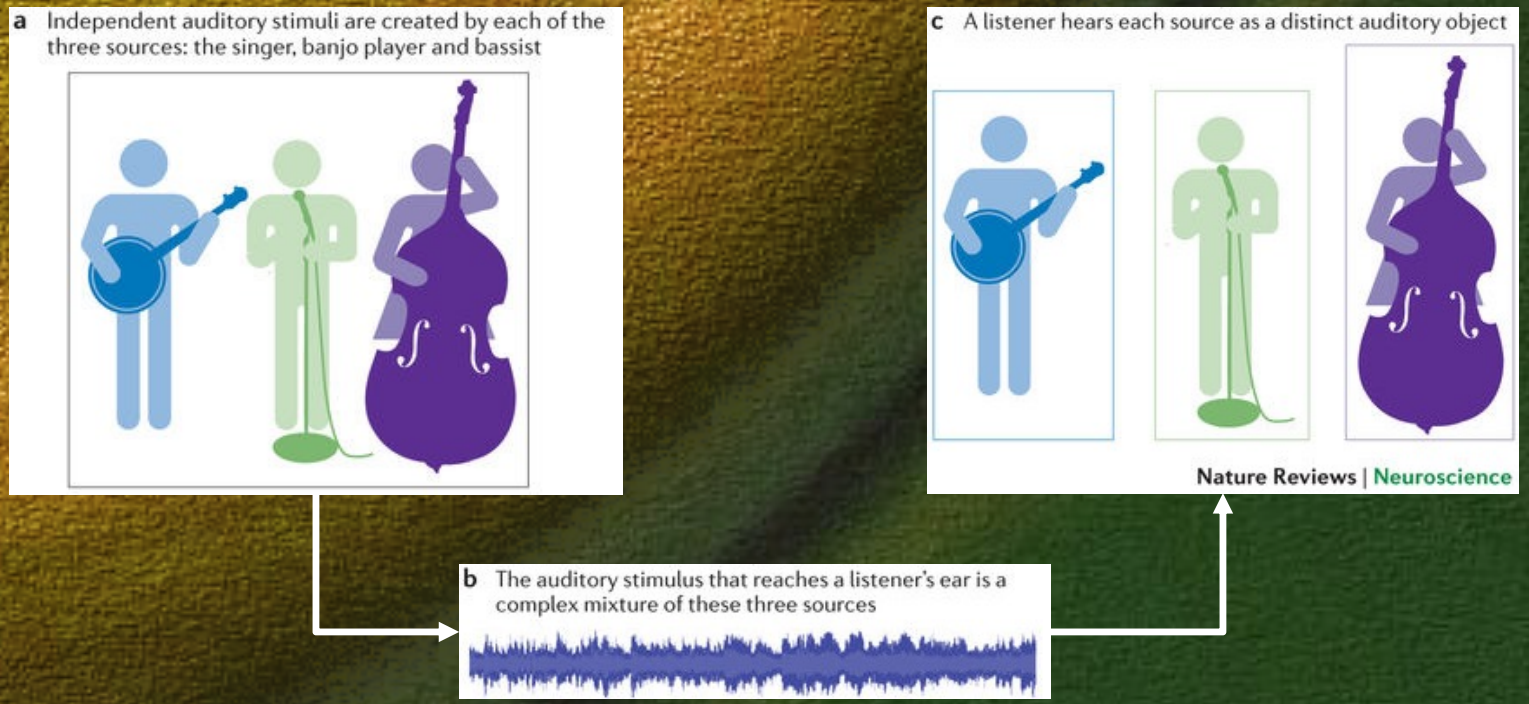
- Al nostro orecchio: unica forma d’onda analizzata nelle sue parziali (informazione grezza)
- Oggetti percettivi complessi da ricombinazione opportuna dell'emissione sonora delle sorgenti
- “re-identificare” le sorgenti: ri-assegnare le parziali alle sorgenti sonore di provenienza

Auditory scene



https://spectrum.ieee.org/consumer-electronics/audiovideo/deep-learning-reinvents-the-hearing-aid, By DeLiang Wang, 2016, Illustration: Emily Cooper

Percezione scena uditiva



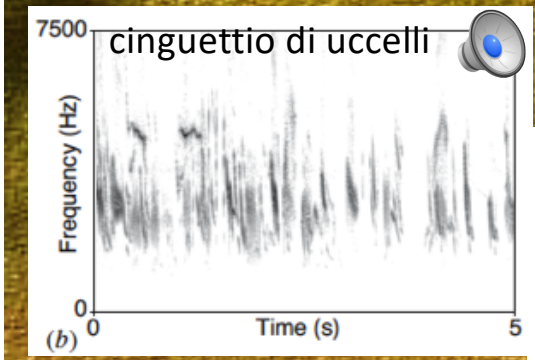
https://www.nature.com/articles/nrn3565,
Jennifer K. Bizley & Yale E. Cohen *Nature Reviews Neuroscience* **volume 14**, pages 693–707 (2013)

Analisi della scena uditiva

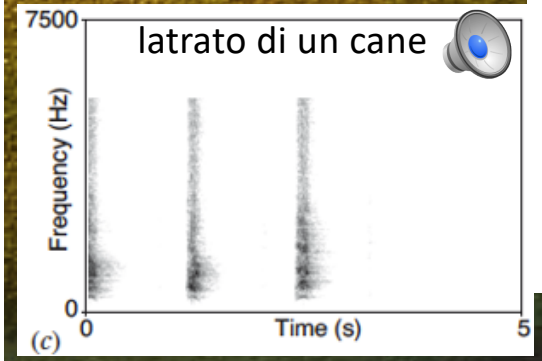
- Scena comprende tutto il percepito: un continuum acustico “grezzo”
- Ricostruzione eventi del mondo che sono causa degli eventi sonori nel continuum
- Figuratività causale: formulazione di una “storia interessante e consistente a proposito del suono”

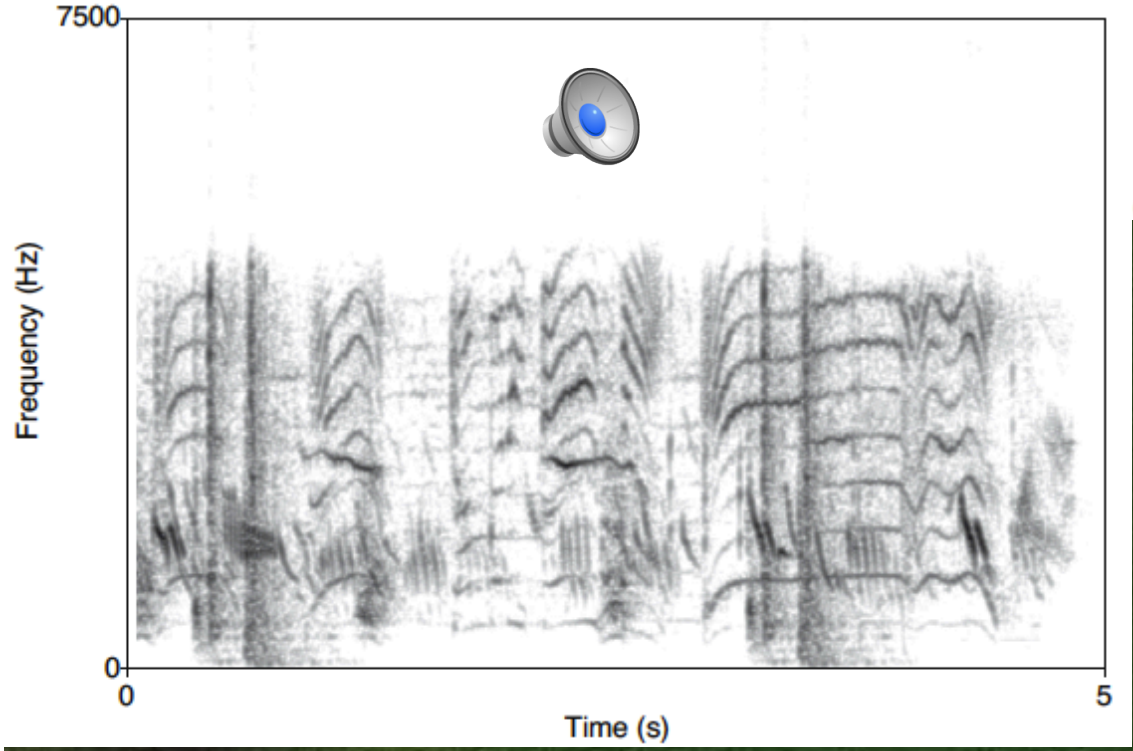
Esempio

cinguettio di uccelli

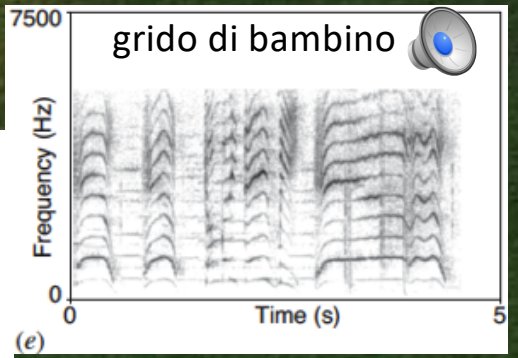


latrato di un cane

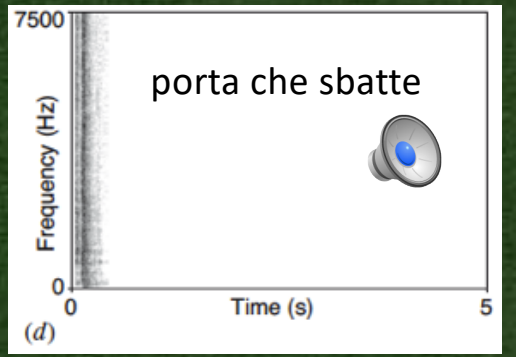




grido di bambino



porta che sbatte



Problema

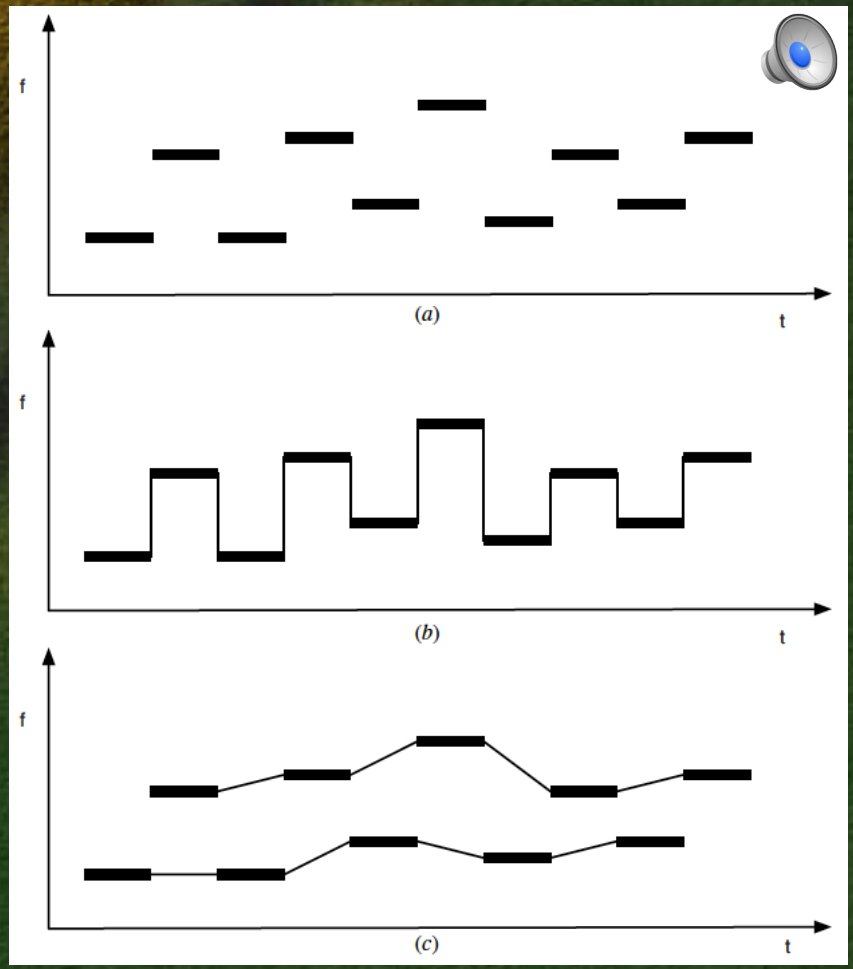
- Generale instabilità degli oggetti uditivi (quanti sono gli strati?)
- Risultato di un complesso lavoro di analisi svolto dal sistema uditivo (quali euristiche?)
- Valutazione dei risultati proposti da euristiche in conflitto

Segnali sinusoidali e auditory stream

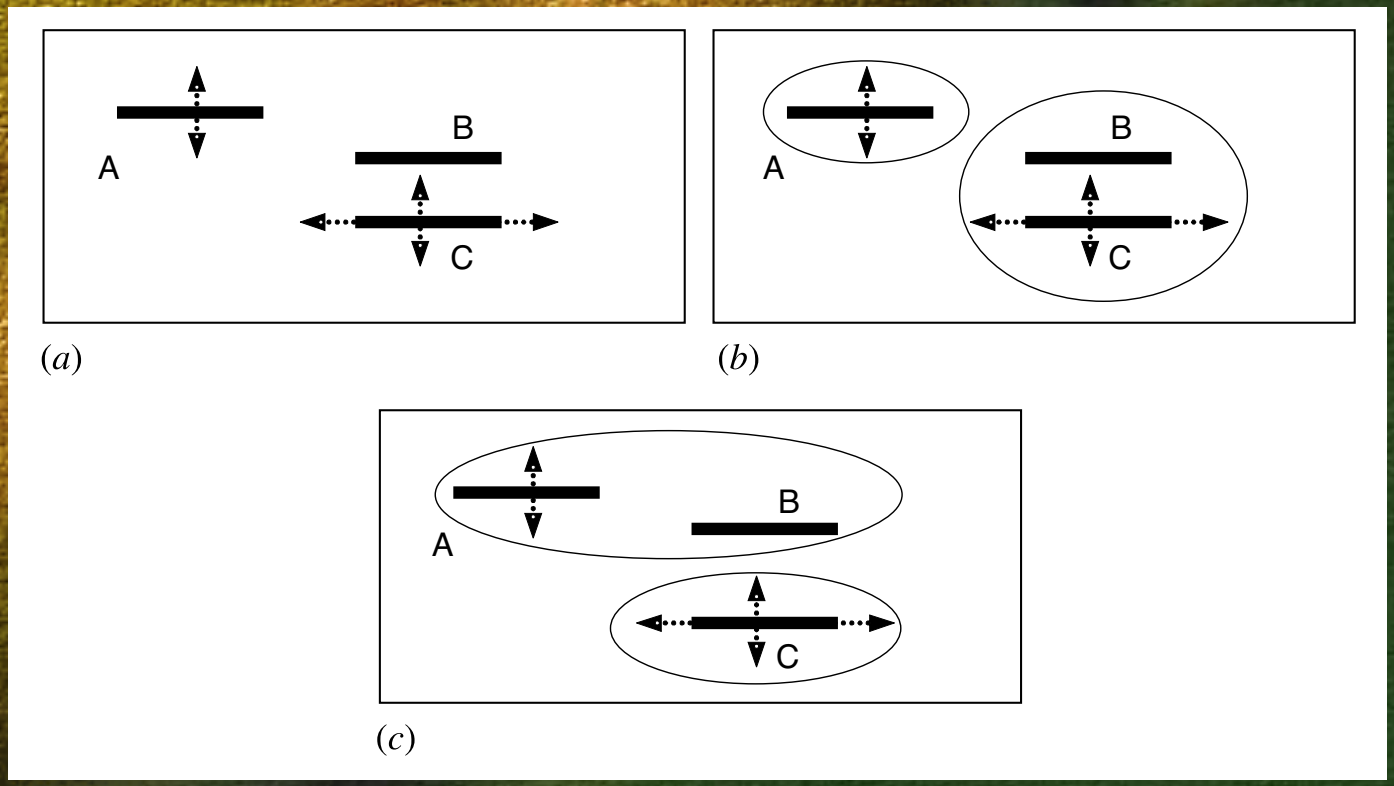
Tendenza a segregare in due flussi più forte per
- maggiore separazione in tessitura
- più elevato tasso di presentazione degli stimoli

1) sequenza come unico flusso,
melodia dal profilo oscillante,
ottenuto raggruppando il sonoro rispetto alla tessitura
(*simultaneous streaming*, "in verticale")

2) due flussi simultanei che
occupano due tessiture differenti
ottenuti raggruppando il sonoro rispetto al tempo
(*sequential streaming*, "in orizzontale")



Scena uditiva minimale



Esercizio: ricreare questa situazione in uno sketch Processing

Euristiche

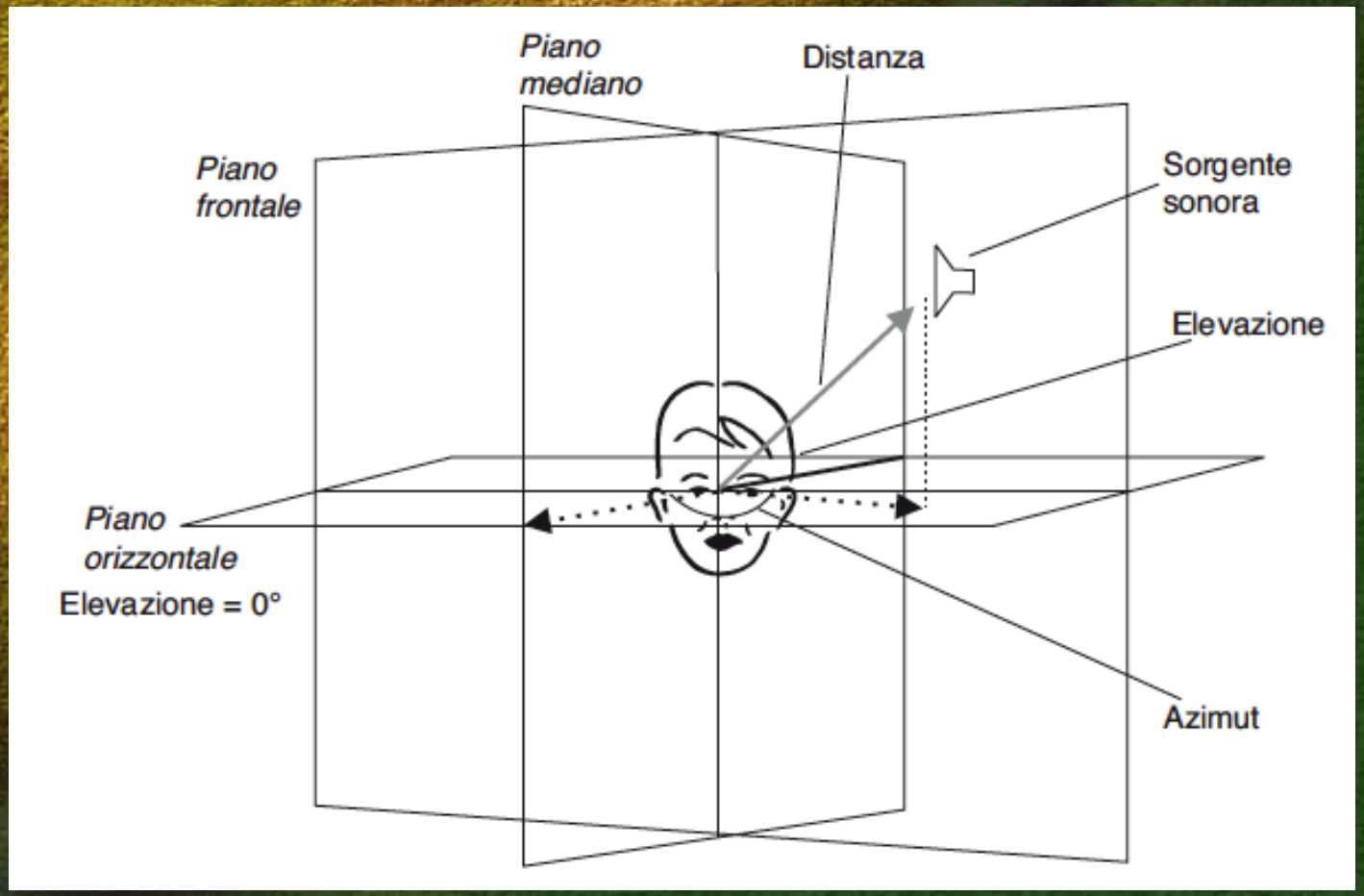
- Euristiche basate su primitive percettive (Gestalt)
  - somiglianza: prossimità di altezza, …
  - buona continuazione in frequenza, …
  - destino comune, delle componenti frequenziali
  - allocazione esclusiva ad un flusso
  - chiusura con effetti di mascheramento
  - salienza, gerarchia tra figura e sfondo
- Euristiche basate su schemi cognitivi appresi
  - non elencabili, dipendono da contesto culturale/personale
  - fogli udibili

Localizzazione dei suoni

Localizzazione delle sorgenti sonore

- Costruzione della mappa sonora soggettiva
  - Localizzazione vera e propria delle sorgenti
  - Caratterizzazione dell'ambiente circostante
- Localizzazione: direzione e distanza
- Ambiente: spazi frequentati senza sorgenti specifiche, "spaciousness"

Posizionamento di sorgente

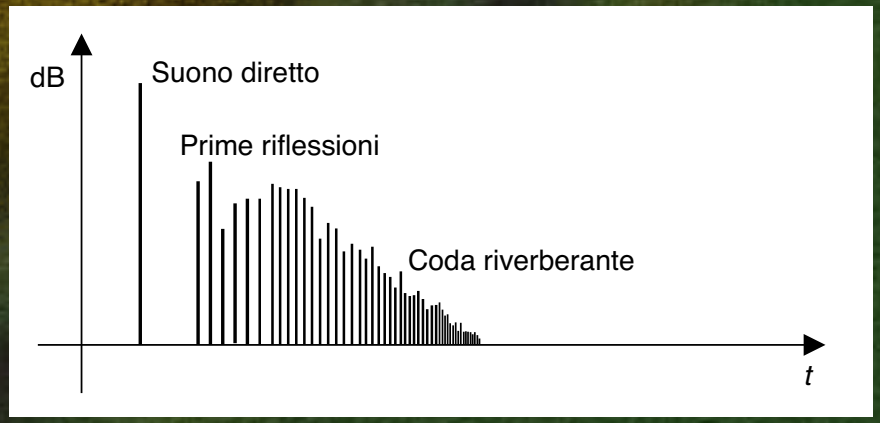


Caratterizzazione ambiente sonoro

- Outdoor / Indoor
  - Assenza/presenza di riflessioni (free field e camera anecoica)
  - Distanza delle sorgenti sonore
  - Dimensione dello spazio: suono diretto / suono indiretto
- Outdoor: facile per direzione, difficile per distanza
- Indoor: difficile per localizzazione, facile per distanza, riflessioni utili per dimensione

Sorgente e ambiente

- Sorgente più o meno direzionale: rapporto tra
  - decibel diffusi nella direzione preferenziale /
  - decibel diffusi in tutte le altre direzioni
- Sorgenti più direzionali con alte frequenze
- Risposta dello spazio



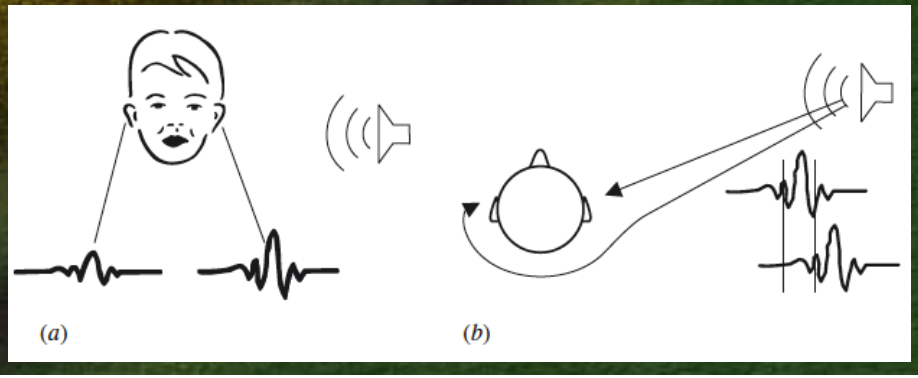
Evidenza sperimentale sul lobo

- Localizzazione monoaurale: interazione tra suono diretto nel canale uditivo e suono riflesso dalle pieghe
- Localizzazione binaurale: filtraggio spettrale operato dal lobo (altezza e posizionamento davanti/dietro rispetto all'ascoltatore)

http://www.ese.wustl.edu/~nehorai/research/biomim/hearing.html

La teoria Duplex (Lord Rayleigh)

- Localizzazione del suono basata su differenze interaurali
  - di intensità alle alte frequenze (IID)
  - di fase alle basse frequenze (ITD)
- Teoria valida per i toni puri o suoni a regime
- Teoria attraente per gli ingegneri del suono

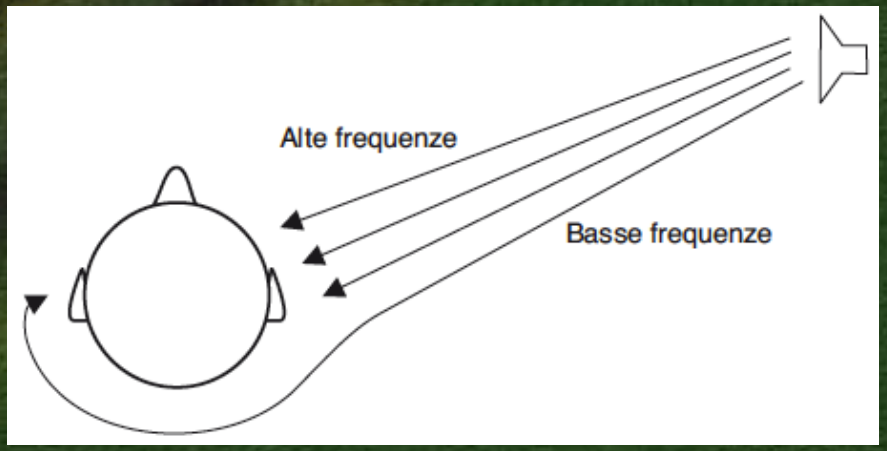


Ruolo della testa

- Alte frequenze:
  - la testa getta “un’ombra acustica” (filtro passa-basso)
  - volume relativo del suono alle due orecchie differente
- Basse frequenze:
  - il suono subisce una diffrazione e avvolge la testa
  - ritardo tra i due suoni

Banda di transizione
intorno a 1500 Hz

λ = 18 cm



Calcolo ITD

Aumento di istanza per orecchio controlaterale

$$r \sin \theta + r \theta$$

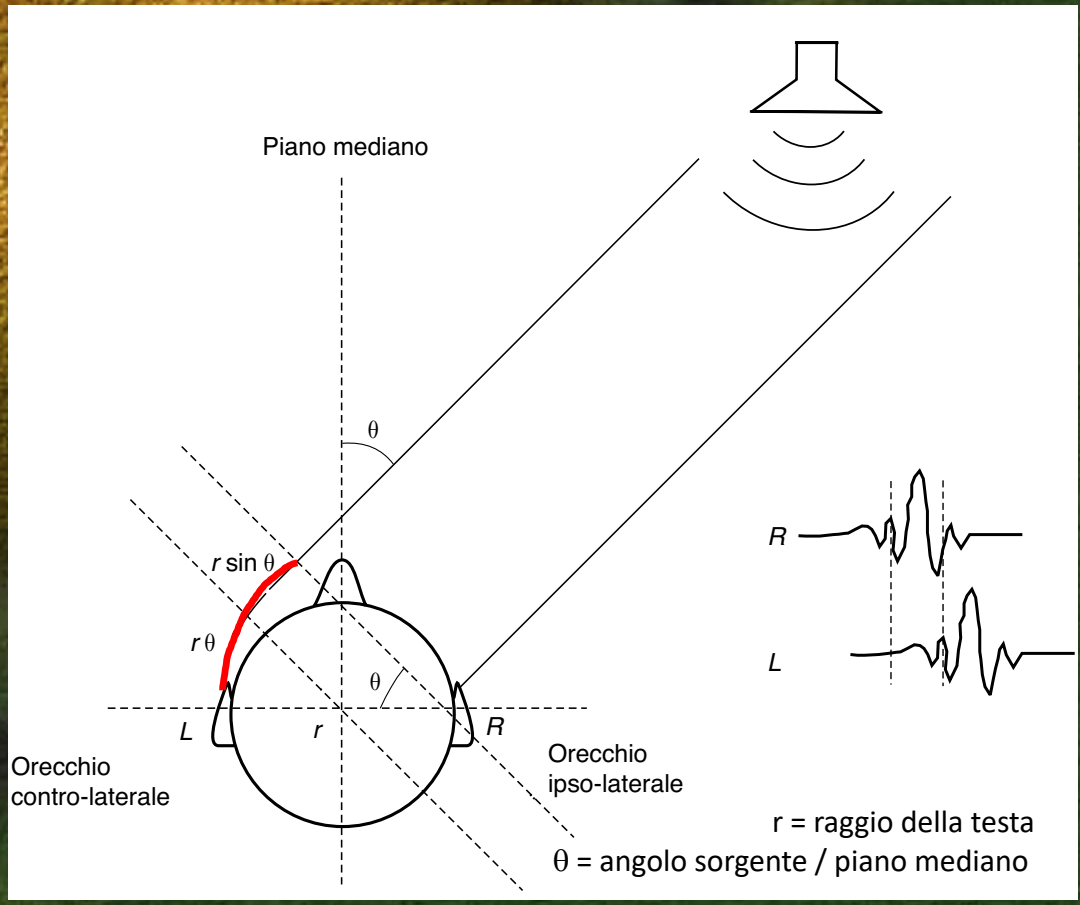
Tempo necessario per questa frazione in più

$$(r \sin \theta + r \theta) / c$$

c = 340 m/s, r = 9 cm

Per $\theta = 90^{\circ} = \pi/2 = 1{,}57$ rad angolo massimo di azimut sul piano orizzontale

9 x 1 + 9 x 1,57 = 23,13 cm
23,13 / 34 ms = 0,65 ms

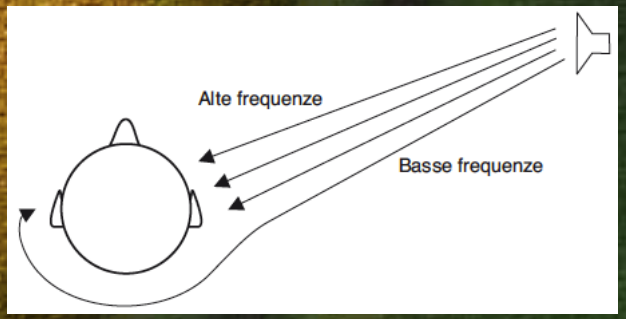


Ruolo ITD

Localizzazione di una sorgente sonora entro 1°→ ITD di 0,01 ms, min ITD rilevabile 0,006 ms

- ITD efficace per i suoni complessi nelle fasi transitorie (attacco e rilascio)
- Si basa sulle basse frequenze
- Discrimina tra sx e dx; no fronte/retro, elevazione
- N.B. sorgenti multiple, effetto di precedenza

Calcolo IID

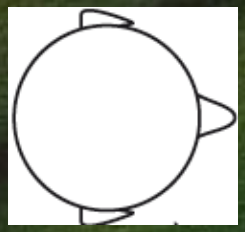


Differenza di ampiezza dovuta a filtraggio spettrale (localizzazione monoaurale)

- Testa
- Padiglione auricolare (risonanza concha)
  - Rilevazione elevazione
  - Rilevazione fronte/retro
- Spalle e corpo

Head Related Transfer Function HRTF

- Funzione di trasferimento in relazione alla testa
- Cambiamenti di forma d’onda, fase e ampiezza
- Sorgente in movimento rispetto all’ascoltatore



Head Related Transfer Function HRTF

- Misurazione con microfoni posti nell'orecchio (dummy head)
  - condizioni di controllo assoluto sull'ambiente
  - differenza tra i segnali alle due orecchie

Risultati misurazioni HRTF

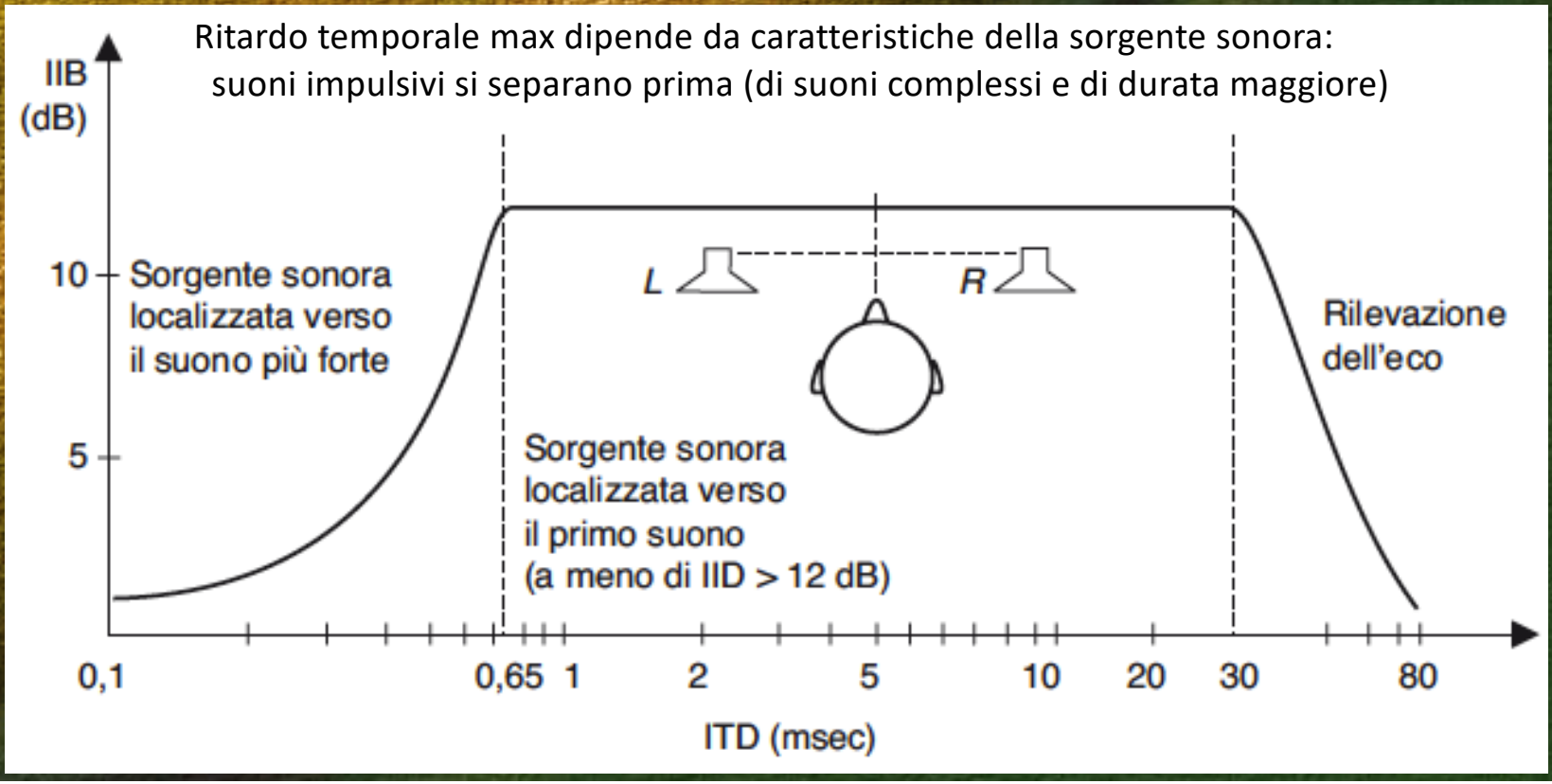
- Sorgenti poste dietro rivelano povero contenuto di alte frequenze (orientamento padiglione)
- Alcune regioni (bande) dello spettro enfatizzate in determinate direzioni
  - intorno a 8 kHz per sorgenti posizionate sopra la testa
  - 300-600 Hz e 3000-6000 Hz per suoni frontali
  - intorno a 1200 e 12.000 Hz per suoni posizionati dietro
- HRTF associate con ITD, per localizzazione, ma difficili da generalizzare

Effetto di precedenza

- Legge del primo fronte d'onda, effetto Haas o legge di soppressione dell'eco
- Scenario con più di una sorgente (due sorgenti simili in posizioni diverse)
- Si percepisce una direzione che corrisponde all'incirca alla prima sorgente (entro certi limiti)
  - Eco molto più forte del primo suono prima di poter essere percepita
  - Aumento di ampiezza deve essere superiore man mano che il ritardo temporale diminuisce

Bilanciamento tra IID e ITD
Effetto anche dopo 0,65 ms

Ritardo temporale max dipende da caratteristiche della sorgente sonora:
suoni impulsivi si separano prima (di suoni complessi e di durata maggiore)

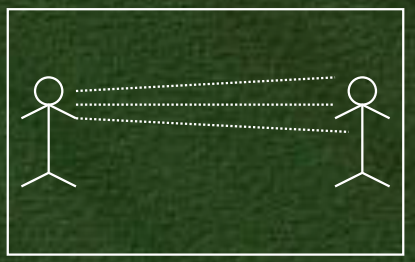


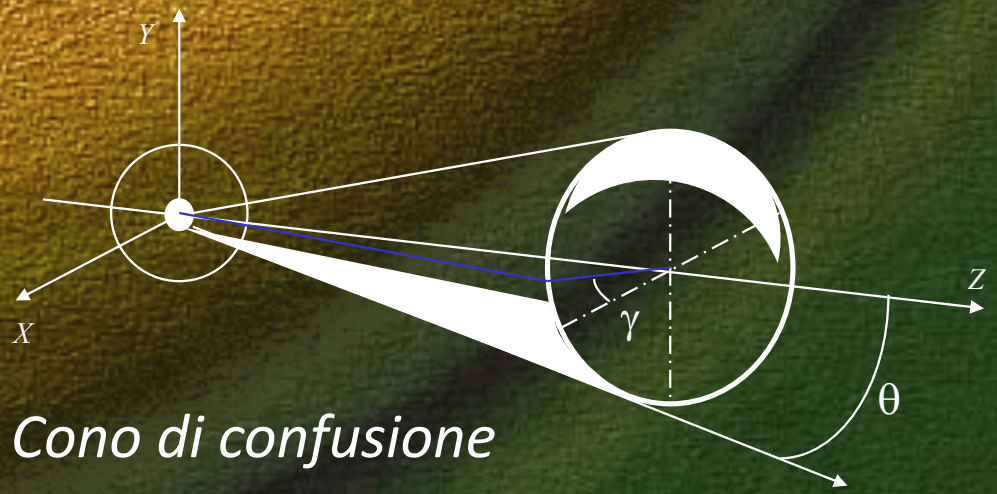
Effetto di precedenza in stereofonia

- In cuffia e altoparlanti
- Localizzazione desiderata variando ITD e IID
  - intensità in dB può compensare alcuni ms di ritardo
  - efficacia dipende dalla natura della sorgente

Il cono di confusione e le applicazioni

La percezione della direzione dipende da almeno 4 fattori complementari





Cono di confusione

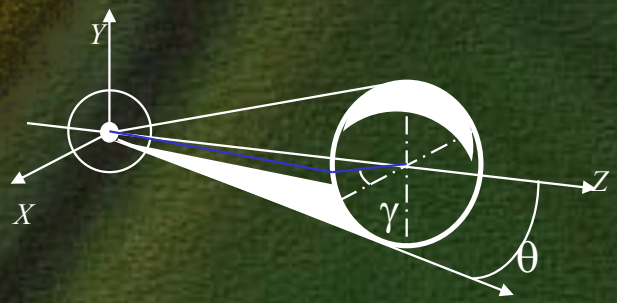
1. Rilevamento IID

- Opera sia con suoni transitori che con suoni a regime
- E’ efficace soprattutto alle alte frequenze (maggiori di 1,5 KHz)

2. Rilevamento ITD

- Funziona bene sotto i 1500 Hz
- Alle alte frequenze
  - cellule nervose non possono scattare tanto velocemente da mantenere l'info di fase
  - metodo ambiguo: alcuni ritardi potrebbero essere più lunghi di un ciclo
- Contribuisce alla lateralizzazione

3. Rilevamento tempi di attacco



- solo per suoni transitori (~100ms)
- distanza tra le due orecchie = ~15cm: il suono viaggia per altri 20 cm (0.65 ms più tardi)
- lateralizzazione del suono entro pochi gradi

4. Forma orecchio esterno

- Usata per distinguere il davanti dal dietro
- Efficienza di convogliamento per le alte frequenze (> 5 KHz) dipende dalla direzione
- Forza relativa differente tra le componenti ad alta frequenza (davanti VS. dietro)

… Il movimento della testa

- Si muove per captare come cambia il suono
- Con un breve suono e testa rigidamente ferma, raramente sicuri della direzione
- Muovendo la testa e/o suono continuo o ripetuto, identificazione accurata

Conclusioni

- Fisiologia dell'orecchio e funzionamento tonotopico della coclea
- Fisica-percezione-cognizione (diagramma di Fletcher-Munson e identificazione sorgenti sonore)
- Interferenza tra i suoni: mascheramento
- Organizzazione percettiva della scena sonora
- Localizzazione delle sorgenti sonore


Grazie dell’attenzione