

Perché Udiamo ciò che udiamo

Giovedì, 22 marzo 2012 – da Stereophile

Tradotto da Mario Bon – n.d.t. = nota del traduttore

Parte prima = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-1.html>

Parte seconda = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-2.html>

Parte terza = <http://audioskeptic.blogspot.it/2012/03/final-steps-in-auditory-system-four-key.html>

Parte quarta = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-4-end.html>

n.d.t. La traduzione non è facile perché il termine loudness non è traducibile e perché l'autore usa il termine "compressione" in luogo di "mascheramento".

Perché Udiamo ciò che udiamo, Parti 1,2,3 e 4

Questa panoramica spiega perché sentiamo ciò che sentiamo e come si può manipolare l'ascolto per ottenere risultati diversi da più ascolti (della stessa cosa).

Background e Note su questo tutorial

Questa guida contiene informazioni ottenute in 26 anni come ricercatore presso i Bell Labs in Acoustic Research a Murray Hill, e dei suoi diretti discendenti, da una varietà di documenti, di quanto ho imparato mentre ero un audio_architetto presso Microsoft, e quando ero Chief Scientist per Neural Audio. L'articolo contiene idee raccolte da una varietà di documenti ed esperimenti fatti da molte persone, su un lungo periodo di tempo.

Le informazioni qui presentate sono un "work in progress" di ricerca nell'interfaccia esterna e con questo intendo l'orecchio, fino al coclea, e del Sistema Nervoso Centrale (SNC), che è il cervello.

n.d.t.

In sostanza l'autore non distingue, come si fa di solito, l'orecchio esterno, medio, interno, il nervo acustico ed il cervello. Fa una semplificazione: interfaccia esterna o periferia (testa, orecchio) e Sistema nervoso centrale (nervo acustico e cervello). Chiama queste regioni "periferie" e considera i fenomeni ed i processi che avvengono nell'una e nell'altra senza scendere troppo nei dettagli (come si fa nella Teoria dei Sistemi). Nel seguito il termine periferia sarà tradotto con interfaccia (nel senso di periferica come quelle dei computer)

Presenterò una panoramica di alto livello di un ampio numero di argomenti. Si tratta di fenomeni osservati sia dai ricercatori in campo dell'udito che in psicologia cognitiva. Ognuno di questi fenomeni proviene da studi approfonditi, alcuni dei quali saranno oggetto di articoli futuri.

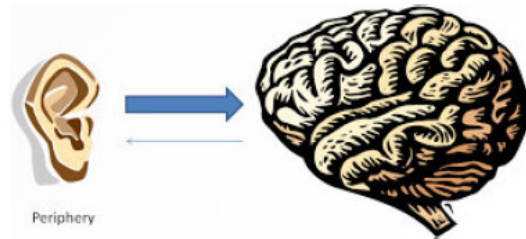
Tenete a mente, queste informazioni (per quanto riguarda l'udito)

- Non è inviolato (n.d.t. nel censo che non è completamente sconosciuto)
- È una discussione di fenomeni (n.d.t. forse intende un "insieme di fenomeni")
- Nella maggior parte dei casi ciò che avviene oltre la membrana basilare è sconosciuto
- Sarà rivisto da ulteriori ricerche col passare del tempo, ci saranno sempre delle revisioni

Una introduzione al sistema uditivo

Il sistema uditivo consiste, nel modo in cui di solito è esaminato, di due parti, la l'interfaccia esterna (testa, orecchio, coclea) ed il cervello o Sistema Nervoso Centrale (SNC). Il flusso di informazioni è sostanzialmente unilaterale, con l'orecchio che fornisce un sacco di informazioni al cervello, e il cervello che fornisce un feedback molto limitato (alcuni problemi di volume, corpo e movimento della testa) verso la l'interfaccia, relativamente parlando.

n.d.t : il 95% delle informazioni che transitano lungo il nervo acustico vanno dall'orecchio al cervello. Il 5% dal cervello all'orecchio (e vanno alle cellule ciliate esterne). Questo feedback "blocca" alcune informazioni ed è quindi molto importante.



Quale è l'interfaccia (esterna)?

È tutto l'apparato dell'udito che si trova all'esterno del cervello. Io non includo, per questo scopo, il corpo e la sensazione della pelle, che sono sensibili, per lo più, a livelli di suoni più alti di quello che dovremmo ascoltare. Dividerò l'interfaccia esterna in tre parti distinte

- HRTF, tra cui il canale uditivo (funzioni dell'orecchio esterno e medio)
- Analisi cocleare (orecchio interno)
- Riduzione del suono in loudnesses parziali in funzione del tempo (orecchio interno)

In primo luogo, la forma fisica della testa, l'orecchio, il corpo e l'ambiente circostante creano quello che viene chiamato Head Related Transfer Function (HRTF) distinte per ciascun orecchio. Si può pensare alla HRTF come una risposta in frequenza che varia con la distanza e l'angolo tra la sorgente sonora e la testa, formando un Interaural Livello Difference (ILD = Differenza di Livello Interaurale) tra ciascun orecchio. La HRTF crea anche le informazioni Interaural Time Delay (ITD = Tempo di Ritardo Interaurale).

L'ITD e ILD sono le informazioni che il sistema nervoso raccoglie dalla realtà fisica acustica per l'elaborazione direzionale, come spiegherò più avanti.

In questo tutorial, tutti gli effetti dovuti all'orecchio esterno e medio sono stati semplificati e aggregati nella HRTF. Convenzionalmente in letteratura, le funzioni dell'orecchio esterno e medio sono separati gli uni dagli altri.

Come contribuisce l'orecchio medio alla HRTF? Il condotto uditivo influenza la HRTF con la sua risonanza, che dipende dalla sua lunghezza e larghezza. Il timpano e 3 ossa dell'orecchio medio influenzano il HRTF modificando l'impedenza del sistema e impedendo il passaggio delle componenti quasi-continue del suono (n.d.t come un filtro passa alto). In altre parole, è simile a un trasformatore che modifica il rapporto tra forza e spostamento (come un filtro passa alto) che elimina le frequenze molto basse quali i cambiamenti nella pressione barometrica. La risonanza da parte dell' orecchio medio dei segnali quasi-continui è assolutamente essenziale: se così non fosse il passaggio di una perturbazione atmosferica equivarrebbe a un livello sonoro di 150dB, intollerabilmente forte e dannoso per il nostro udito.

In secondo luogo, la coclea filtra il segnale in molte bande parzialmente sovrapposte. La coclea è un organo complicato che esegue un filtraggio meccanico del segnale che entra dall'orecchio. Filtra il segnale in ingresso in bande critiche fortemente sovrapposte o larghezze di banda rettangolari equivalenti (ERB). Questo filtraggio funziona per frequenze superiori a circa 100Hz.

n.d.t. queste bande, dette bande critiche, sono 24 e, a parte quelle a bassa frequenza hanno la larghezza di circa un terzo di ottava. Sotto i 100Hz le bande sono più larghe tanto che la distorsione di seconda armonica cade all'interno della stessa banda che contiene la fondamentale e viene mascherata dalla fondamentale stessa. Per questo motivo l'orecchio è poco sensibile alla distorsione armonica a bassa frequenza: non la sente. Invece la distorsione di terza armonica e quella per intermodulazione la sente benissimo.

In terzo luogo, il segnale in ogni banda ERB viene compresso per darci una riduzione del suono in loudnesses parziali in funzione del tempo.

All'interno della stessa banda critica (della stessa ERB) i suoni a livello più alto mascherano quelli a livello più basso. L'informazione quindi si semplifica. In questo senso il segnale viene "compresso" nel senso che vengono eliminate delle informazioni.

Se le cose andassero effettivamente così ascoltare musica "compressa" in mp3 dovrebbe essere meno affaticante che ascoltare la musica non compressa mentre l'ascolto della musica in alta definizione dovrebbe essere molto affaticante: infatti il cervello lavorerebbe meno con l'mp3. Anzi più l'mp3 è compresso e meno faticoso dovrebbe essere l'ascolto. Dato che ciò non succede, e l'alta risoluzione "suona meglio dell'mp3) significa che il meccanismo non è esattamente quello descritto.

Questa lega i cambiamenti della pressione atmosferica alla sensazione. Un livello di pressione sonora (SPL) è una grandezza fisica ovvero una variazione della pressione atmosferica. Una loudness parziale è la misura

del livello di sensazione rilevato dalle cellule dell'orecchio interno, misura una grandezza in percentuale. Un aumento dell'intensità di 10dB crea un aumento della sensazione di un fattore 2.

Se si chiede ad un individuo di aumentare il volume del doppio, questi lo aumenterà di 10 dB. Questo non significa che la minima variazione percepibile di livello si di 10dB. Sappiamo che, con rumore rosa e su larghezza di banda abbastanza ampia, è possibile discriminare anche una frazione di dB. Secondo alcuni 0.1 dB che corrisponde ad una variazione di ampiezza dell'1%. Se ciò fosse vero significherebbe poter percepire la differenza tra un rumore rosa riprodotto con un tasso di distorsione dello 0.1% dallo stesso rumore riprodotto con un tasso di distorsione dell'1%.

Un segnale su una ERB non maschera i segnali che non entrano nella stessa banda critica, così ci ritroviamo con un effetto interessante: i segnali sono mascherati all'interno di una ERB, ma le loudness parziali di 2 diverse Erbs si sommano! La somma di tutti le loudness parziali è in realtà quello che noi percepiamo come loudness.

n.d.t. in realtà la somma va fatta seguendo una regola abbastanza incasinata che si deve a Zwicker http://medi.uni-oldenburg.de/download/docs/diss/Launer_1995_LoudnessPerception/appendix.pdf

Si prega di notare che ho usato due parole in un modo molto specifico. Per essere chiari,

- Loudness è il livello che sentite, il livello di sensazione che viene spedito al CNS (al cervello)
- Intensità è il livello misurato del segnale misurato da un microfono, l' SPL

n.d.t

in realtà l'intensità è una cosa e il livello un'altra. In particolare il livello è per definizione il logaritmo del rapporto di due grandezze omogenee (di cui una presa per riferimento). L'intensità acustica è quella quantità che, integrata su una superficie sferica, fornisce la potenza acustica quindi è potenza acustica per unità di superficie. Si capisce quindi perché queste due grandezze non si possano corrispondere.

Le due non corrispondono molto bene, anche tenendo conto degli effetti di mascheramento di cui sopra. L'unico caso in cui si tracciano (si corrispondono) è quando la risposta in frequenza dei due segnali a confronto hanno la stessa forma e differiscono solo per l'energia totale. Una misura di SPL non è sufficiente per determinare il loudness (n.d.t. ci mancherebbe solo questo!!!).

Per ulteriori informazioni sul loudness, consultare il tutorial Pacific Northwest AES web site (Loudness Tutorial PowerPoint presentations without sound, with audio).

Ci sono una quantità di diapositive sui temi di audio e dell'udito, che si potrete trovare interessanti.

Qual è il sistema nervoso centrale?

Il sistema nervoso centrale (SNC) inizia all'estremità del nervo uditivo dove si collega al cervello e comprende poi il cervello. Svolge, in qualche modo che non abbiamo ben capito, le seguenti operazioni:

- Riduzione da loudness parziali a qualità uditive
- Riduzione da qualità uditive a oggetti acustici (n.d.t. "estrazione dei contorni" di un suono)
- Memorizzazione a breve e lungo termine

Ci sono alcuni problemi noti con il Sistema Nervoso Centrale. In particolare, è molto flessibile nel modo di interpretare gli stimoli. Questo è talvolta indicato come "plasticità".

n.d.t. : di norma il termine plasticità è riferita al cervello ed indica la capacità, del cervello, di modificare le connessioni tra i neuroni. La plasticità decade con l'età per cui gli anziani fanno più fatica ad apprendere e a cambiare idee ed abitudini.

Plasticità significa che può cambiare ciò che si sente, ciò che si vede, o l'odore ..., e il sistema nervoso centrale, per come è fatto, combina le informazioni provenienti da tutti i sensi. Lo fa sempre ed in tutte le circostanze. Ovunque. Tutto il tempo.

La plasticità, di per sé, crea un problema nell'isolare quello che sta succedendo in qualsiasi bit (spezzone) di audio, sia esso dovuto alle apparecchiature, prestazioni, o quello che avete, perché la sola conoscenza di ciò che state usando per ascoltare cambierà il modo di ascoltare e quindi cambierà l'oggetto della vostra attenzione.

Questa non è una questione sul tipo della incertezza di Heisenberg, piuttosto si tratta di un semplice caso in cui si notano le cose che si osservano (su cui ci si concentra).

Quali informazioni ottiene al SNC e quando ci arriva (dagli stimoli esterni)? Tutto quanto rilevato dall'interfaccia uditiva, la vista, la conoscenza del bottone che avete pigiato sull'amplificatore, il colore del tessuto della griglia dell'altoparlante, e così via ... arriva al SNC. Tuttavia, ciò che è più importante per la sensazione uditiva sono le informazioni che provengono dalla interfaccia uditiva (dall'orecchio).

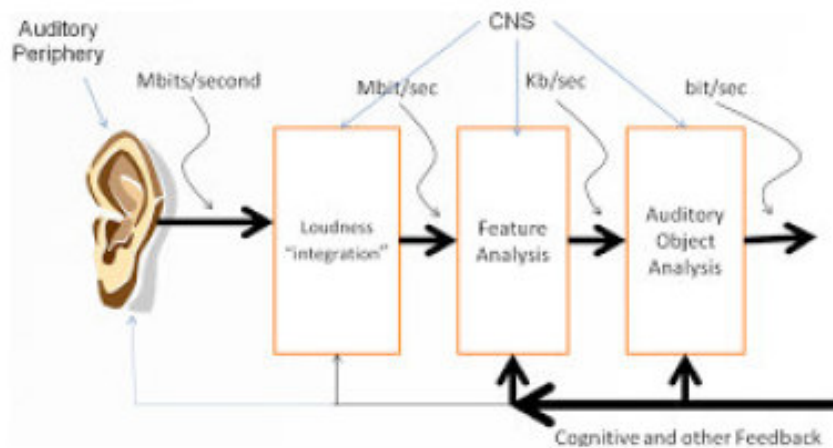
Qui, non stiamo considerando le sensazione del corpo (che riguardano le basse frequenze ad alti livelli, o le frequenze ultrasoniche a livelli estremi). Lasciamo da parte le basse frequenze estreme di intensità estrema ed i segnali VHF (Frequenze Molto Alte) che possono essere rilevati con altri mezzi; queste sono condizioni estreme e non dovrebbero generalmente essere sperimentate da un ascoltatore.

n.d.t. l'autore si limita a descrivere le "effettive condizioni d'uso".

Le frequenze sotto i 10 Hz s SPL di 100 dB e oltre stimolano il rilascio di ormoni, vomito, nausea e cefalea.

Gli ultrasuoni appena oltre 1 20kHz e livello sufficiente eccitano direttamente i neuroni

Per riassumere, il CNS e la sua connessione alla interfaccia audio possono essere illustrate come



Quattro punti chiave

Qui ci sono quattro punti chiave che riguardano la interfaccia uditiva.

- La interfaccia uditiva analizza tutti i segnali in blocchi tempo/frequenza chiamati ERB o Bark. (il Bark è l'unità di misura dell'altezza soggettiva e corrisponde a 100 Mel, la banda audio, composta da 24 bande critiche corrisponde esattamente a 24 Bark, infatti essa è legata alla frequenza M centrale... n.d.t.)
- Grazie alla meccanica della coclea, primi arrivi sono molto forti, con influenza apparentemente sproporzionata su ciò che realmente si sente. Questo è effettivamente utile nel mondo reale.
- I segnali all'interno di una ERB si mascherano reciprocamente (n.d.t. reciprocamente mica tanto: esiste una frequenza mascherante e una mascherata).
- Segnali di fuori di un ERB non si mascherano

La prima informazione in arrivo in una data ERB, dove il primo arrivo è il primo segnale in circa gli ultimi 200 millisecondi, è enfatizzata dalla meccanica della coclea. Il mascheramento dei segnali all'interno di una ERB inizia ad avere effetto circa 1 millisecondo dopo l'arrivo del suono, quindi la prima parte di un suono invia più informazioni al cervello. Questo risulta essere molto utile per noi per distinguere entrambe le direzioni di provenienza del suono percepite e la sensazione di diffusione (del suono).

I loudness parziali dalla coclea sono integrati da qualche parte al confine del sistema nervoso centrale in modo tale che qualche ricordo del passato è mantenuto per un massimo di 200 millisecondi. Gli esperimento di Level Roving dimostrano che quando esistono ritardi che si avvicinano a 200 millisecondi tra le due sorgenti, la capacità di discernere le differenze sottili in volume o il timbro sono ridotte. **È ben noto che è necessaria una commutazione molto veloce (e priva di click) tra due sorgenti quando si cerca di individuare piccole differenze, altrimenti si perde parte della capacità di distinguere le differenze di loudness.**

Passaggi finali

Ci sono due passaggi rimanenti per giungere a quello che si percepisce (al suono cosciente), entrambi eseguiti dal cervello in modo che, in questo momento, non interessa capire nemmeno approssimativamente. Il primo passo è l'analisi delle loudness (n.d.t. del contenuto delle 24 bande critiche)

Le loudness parziali vengono integrate, in qualche modo, in quello che io chiamo "**caratteristiche uditive**". C'è una grande quantità di perdita di dati in questo processo, solo un millesimo delle informazioni presenti al nervo uditivo rimane dopo la riduzione delle "caratteristiche uditive", il resto viene integrato (sommato) o scartato. A questo livello, l'analisi è estremamente plastica e può essere guidata dall'apprendimento, esperienza, conoscenza, riflesso, stimoli visivi, stato d'animo, il comfort, e tutti gli altri fattori. Le caratteristiche durano, nella memoria a breve termine, per pochi secondi (n.d.t. chi dice due, chi dice tre secondi).

Nella seconda fase, le informazioni vengono ulteriormente ridotte di circa 100 fino a 1000 volte, e diventano ciò a cui mi riferisco come "**oggetti uditivi**". Queste sono cose che si possono descrivere consapevolmente descrivere, controllare, interpretare, ecc. Le parole sono esempi di cose fatte con una successione di oggetti uditivi.

n.d.t. in realtà le singole parole del linguaggio vengono trattate come oggetti singoli

Questo processo è plastico quanto plastico può essere, vi potete reindirizzare consapevolmente, essere indirizzato da stimoli visivi, congetture, stimoli inconsci, tra cui la casualità, ed è l'ultimo passo prima che la percezione uditiva possa essere inviato alla memoria a lungo termine.

n.d.t. una delle cose più difficili per il cervello e produrre processi casuali. Tutti i ricordi, infatti, sono tra loro collegati. Se esistesse un ricordo "isolato" non sarebbe raggiungibile (quindi come se non ci fosse). L'Alzheimer, troncando le connessioni, isola i ricordi e li rende non più disponibili.

Si parla di casualità quando le cause sono in numero talmente elevato che non se ne può tenere conto.

In tali condizioni valgono le leggi della statistica. Si veda il Teorema del Limite Centrale.

È interessante notare che questo processo può far emergere una proprietà di oggetto (acustico) se ci si concentra consapevolmente su quella caratteristica (n.d.r. come dire che udiamo quello che desideriamo udire). È questo processo è più influenzato da ogni altro stimolo, compresi quelli generati internamente. È in questi ultimi due passaggi, dove il loudness a breve termine è ridotto a proprietà e quindi ad oggetti, dove integriamo gli stimoli dai nostri sensi e dalla nostra conoscenza, indipendentemente dalla nostra volontà o situazione (n.d.t. questa è la percezione ovvero il processo che carica la sensazione con le memorie pregresse e anche con le sensazioni provenienti dagli altri sensi).

Che cosa significa questo circa ciò che si sente

La plasticità nel formare caratteristiche uditive e oggetti uditivi ha alcune implicazioni molto forti. Se si ascoltano cose diverse, si sentono cose diverse. Questo non è illusione, non è confusione, non è inganno, non è allucinazioni, è semplicemente il funzionamento del cervello. Questo ha un risvolto particolarmente importante per gli appassionati di audio, l'aspettativa farà in modo di sentire le (stesse) cose in modo diverso. L'effetto dell'aspettativa non è sempre positivo, o negativo, per esempio, si potrebbe non sentire quello che ci si aspetta di sentire, ma le aspettative non può essere consapevolmente esclusa dalla vostra esperienza di udito (non ora, non mai).

L'effetto dell'attesa è il motivo per cui il test audio deve essere condotto in "doppio cieco".

(n.d.t. spiega cosa è il doppio cieco ma non si capisce molto...meglio leggere da qualche altra parte.)

Molti lavori hanno dimostrato che stimoli che provengono da uno specifico sperimentatore inducono a rispondere in modo diverso. È interessante notare si può non essere d'accordo con lo sperimentatore in un test in singolo cieco, si può prendere l'altra risposta rispetto a quella preferita dallo sperimentatore, ma le vostre prestazioni saranno influenzate.

Questo può essere dimostrato con il vecchio "backward masking" (mascheramento a ritroso) che è stato originariamente definito in un contesto diverso, in cui è stato affermato che si potevano udire contenuti satanici nella canzone "Stairway to Heaven" quando riprodotta al contrario. È interessante notare che, quando il brano viene fatto ascoltare al contrario ad un pubblico ignaro (questo è successo più volte durante delle lezioni) il pubblico non sente nulla o pochissimi sillabe casuali. Tuttavia quando le "parole" sono prima fatte ascoltare, il pubblico le distingue chiaramente e correttamente, anche se il suono reale presentato non è stato modificato (n.d.t. l'attesa di udire le parole fa udire le parole). Questo è un risultato diretto del nostro modo di capire un discorso, quando ci aspettiamo un discorso con qualche contenuto particolare, guidiamo l'estrazione delle caratteristiche e la risoluzione degli oggetti (acustici) per trovare i componenti del nostro

ambiente audio al fine di estrarre le informazioni, e farlo anche se queste informazioni sono assenti (n.d.t. ancora una volta udiamo ciò che vogliamo udire anche a livello di contesto). Questo tipo di effetto è richiamato nelle "EVP" in cui suoni, sussurri, ecc, vengono riconosciuti nel rumore di fondo delle radio (n.d.t. la voce dei morti, i messaggi alieni...).

In sintesi, la trasformazione effettuata nel cervello è eccezionalmente plastica, e può essere guidata da una varietà di cose. Il risultato di tutto questo trattamento è quello che sentiamo coscientemente (il suono cosciente)

- Se si ascolta qualcosa in modo diverso (per diverse caratteristiche o oggetti)
- Vi ricorderete cose diverse
- Questo non è un'illusione
- Se avete motivo di ritenere le cose possano essere diverse
- È molto probabile che ascoltiate in modo diverso
- Pertanto, vi ricorderete cose diverse

In breve, se si vuole essere sicuri di quello che si sente solo a causa degli stimoli uditivi, è necessario organizzare un test in cieco di tipo appropriato. Tali test non sono né facile né semplici da organizzare, ma sono il modo migliore, e forse il solo, per evitare involontari autoinganni.

Fine