



ITALIAN UNIVERSITY LINE

EDITING MULTIMEDIALE

**Brevi note tecniche sull'audio
digitale con riferimento all'uso
di Audacity**

Docente:
Andreas R. FORMICONI

6 gennaio 2015

1 Introduzione

Fermo restando l'orientamento nettamente pratico dell'insegnamento di Editing Multimediale, con queste note cerchiamo di fornire un minimo inquadramento scientifico. Tutto sommato quanto basta per dare un nome alle cose, non molto di più.

2 Campionamento digitale e frequenza di campionamento

Dal punto di vista fisico i suoni che udiamo sono onde di pressione dell'aria. Tutti sanno che stando sulla terra siamo sottoposti ad una certa pressione atmosferica che - in condizioni di tempo stabile al livello del mare - è pari a 1033 mbar. Questa pressione non è costante. Da un lato è sottoposta a variazioni relativamente lente, dovute alle mutazioni metereologiche o alle variazioni di altezza - anche se prendiamo quota con un aeroplano o ci gettiamo con un paracadute le variazioni si estendono sull'arco di qualche minuto. Dall'altro la pressione è continuamente perturbata dai suoni, che sono caratterizzati da variazioni pressorie molto più veloci. Aprite Audacity e registrate un breve brano, anche solo un secondo. Poi applicatevi più volte lo zoom (**Ctrl-1** per zoomare, **Ctrl-3** per tornare indietro, **Ctrl-2** se ci si perde). Vedrete che le oscillazioni sonore sono ben più piccole di un secondo. La curva che si vede in ogni traccia di audacity è la *forma d'onda* che riproduce l'andamento delle onde di pressione sonore. Non chiudete Audacity, il brano che avete appena acquisito ci servirà fra un momento.

Le onde di pressione vengono trasformate in oscillazioni di altro tipo mediante dispositivi trasduttori: per esempio un microfono trasforma le oscillazioni di pressioni in oscillazioni di un segnale elettrico: quest'ultimo varia in analogia alle variazioni delle onde di pressione. Nei vecchi LP, le ondulazioni del solco variano in analogia con le onde di pressione. Questi sono i cosiddetti segnali analogici. I sistemi che sfruttano i segnali analogici per riprodurre il suono sono molto fedeli ma possono avere il problema di essere rumorosi. Ogni volta che il segnale viene trasformato da un mezzo all'altro si introduce una certa quantità di rumore.

Con il metodo digitale si procede in modo diverso: per *acquisire* un segnale questo viene misurato regolarmente a brevissimi e costanti intervalli di tempo, il segnale viene cioè *campionato*. Perché la cosa funzioni la frequenza di campionamento deve essere molto alta, più alta di quella suggerita dalle variazioni della forma d'onda. Esistono delle condizioni ben precise che garantiscono la capacità di ricostruire tutta la forma d'onda a partire dal suo campionamento.

Questo dovrà essere tanto più fitto quanto rapide sono le variazioni della forma d'onda. L'entità di tale requisito è specificata dal teorema del campionamento di Shannon, il quale sancisce che una forma d'onda può essere ricostruita completamente se viene campionata ad una frequenza che sia almeno il doppio della propria. Se tale criterio non viene soddisfatto, si possono avere delle distorsioni del suono. Come si può facilmente immaginare, le forme d'onda che vediamo in Audacity sono forme campionate e non riprodotte per analogia, dal momento che è un computer digitale quello che ve le mostra. Riprendete ora il brano di prima e insistete - se non l'avete già fatto - a zoomare (**Ctrl-1**). Prima o poi dovrete vedere dei puntini: quelli sono i singoli campionamenti. Se esagerate vedete che i punti distano l'uno dall'altro per qualcosa più di 0.00002 secondi, per la precisione 0.0000226757 secondi. Se fate $1/0.0000226757$ ottenete 44100 punti al secondo: 44100 Hertz ¹.

Noi udiamo suoni in un intervallo di frequenza fra 20 Hz e 20000 Hz. Di conseguenza, secondo il teorema del campionamento dovremmo campionare il segnale sonoro almeno a 40000 Hz (40000 volte il secondo). In effetti

¹L'unità di misura della frequenza è l'Hertz. La frequenza non è altro che il numero di volte che un fenomeno periodico si ripete ogni secondo, per esempio oscillazioni al secondo per un'onda regolare. Un Hertz significa una oscillazione al secondo. L'abbreviazione di Hertz è *Hz*.

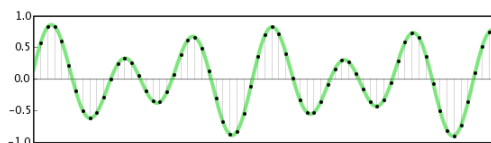


Figura 1: La linea verde rappresenta la forma d'onda, i punti neri sono i valori di essa che vengono effettivamente misurati. Questo metodo si chiama campionamento, ciò che si ottiene in questo processo è esclusivamente l'insieme di valori numerici corrispondenti alle altezze (ordinate) dei punti neri. La figura è tratta dal rispettivo tutorial di Audacity.

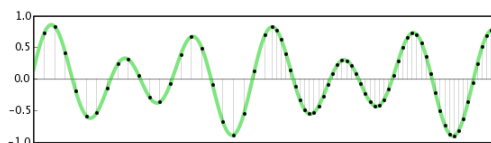


Figura 2: La figura visualizza due diverse frequenze di campionamento, rendendo intuitivo il fatto che, se questa non è adeguata, le variazioni più repentine rischiano di non essere adeguatamente rappresentate. La figura è tratta dal rispettivo tutorial di Audacity.

abbiamo appena trovato che Audacity ha campionato la nostra voce a 44100 Hz, che è il valore più comune, proprio per questa motivazione. In realtà sono usate anche altre frequenze, minori e maggiori. Le più tipiche sono 8 kHz, 16 kHz, 22.05 kHz, 44.1 kHz, 48 kHz, 96 kHz and 192 kHz, dove abbiamo usato l'unità KHz = 1000 Hz. Il parlato è comprensibile anche con frequenze inferiori a 44.1 KHz. Per esempio le linee telefoniche trasmettono segnali compresi fra 200 Hz e 4000 Hz, quindi, sempre per lo stesso criterio, un campionamento di 8 KHz è sufficiente in questo caso.

In conclusione, la frequenza di campionamento determina la qualità del suono campionato: la riproduzione di una sonata per violino, dove tutto si gioca sull'accurata rappresentazione del timbro dello strumento, presenterà requisiti diversi dalla riproduzione di una conversazione telefonica.

3 La rappresentazione dei campioni

L'altro parametro che può influire sulla qualità della riproduzione digitale è il formato dei singoli campioni.

Il campionamento è costituito da una serie di misure. Ognuna di esse è un campione. I campioni possono essere pensati come i pixel delle immagini. In ambedue i casi, ogni campione temporale (campione spaziale per le immagini), consiste in un numero che deve essere messo in memoria. Nella memoria digitale i numeri si rappresentano come numeri binari ². Occorre quindi decidere quanti bit usare per memorizzare le quantità in gioco. Per capirsi, facciamo un esempio stupido: Supponiamo di usare un solo bit. In questo caso potrò rappresentare solo numeri 0 e 1. Proviamo ad usare due bit: 00, 01, 10 11, che poi corrispondono a 0, 1, 2, 3 del sistema decimale. Insomma 4 numeri. Con tre bit: 000, 001, 010 011, 100, 101, 110 111 che corrispondono a 0, 1, 2, 3, 4, 5, 6, 7 del sistema decimale; così con tre bit posso rappresentare 8 numeri. Si capisce facilmente che ogni volta che aggiungo 1 bit raddoppio i numeri che posso rappresentare.

Più sono i bit che io uso per rappresentare il risultato di una misura e maggiore sarà la precisione con la quale io la descrivo: con 2 bit ho solo 4 possibilità, con 10 bit ne ho 1024. La *profondità* in bit determina l'*intervallo dinamico*, o *range dinamico* (da *dynamic range*) del segnale sonoro, ovvero il rapporto fra il massimo e il minimo valore possibile.

²L'elettronica è fatta così, perché questo si è rivelato il metodo in grado di dare le maggiori prestazioni in relazione ai costi.

Il numero di bit dipende da cosa si misura. In questo caso si tratta di suoni, ma soprattutto dei suoni che noi siamo in grado di udire. Ebbene, la capacità del nostro orecchio è mostruosa. E qui entrano in gioco i *decibel*, che si indicano solitamente con il simbolo *Db*

La nozione di *decibel* è tecnica ma compare abbastanza frequentemente nel linguaggio comune. Nella tabella 1 sono mostrati i valori dell'intensità del suono in vari ambienti sonori.

Vediamo dunque cosa sono i decibel.

1. Il decibel esprime l'entità di una grandezza fisica attraverso il rapporto fra il suo valore e un valore di base preso a riferimento. Trattandosi di un *rapporto* è una quantità *adimensionale*. Esempio: se devo suddividere 10 Kg di mele in porzioni di 0.5 Kg ciascuna, faccio $10/0.5$ che dà 20 - questo è un numero puro, non ha più la dimensione di mele.
2. In realtà il decibel non è proprio il rapporto ma il suo logaritmo. Questo è possibile perché il rapporto è un numero puro e le funzioni matematiche si applicano ai numeri puri.

Ma perché il logaritmo? Il motivo sta proprio nella smisurata risposta dinamica del nostro orecchio: l'intensità del suono di una motosega supera di un milione di volte quella dei minimi suoni che siamo in grado di percepire! Poiché lavorare con numeri che variano in domini di grande estensione può generare dei problemi, si usano degli accorgimenti per semplificarsi la vita. Un esempio può essere la notazione scientifica ³, un altro, appunto la rappresentazione logaritmica. Quest'ultima è utile per condensare un grande ambito di variabilità in uno di minore ampiezza, un effetto che si ottiene grazie alla proprietà della funzione matematica *logaritmo*.

Le funzioni matematiche sono dispositivi di viaggio: ti portano da un mondo a un altro, dalle ascisse (asse orizzontale delle x) alle ordinate (asse

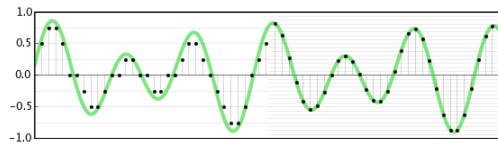


Figura 3: La metà destra della traccia è stata campionata con soli due bit per misura, per cui ognuna di esse può assumere solo 4 valori - si apprezza come i reali valori che starebbero sulla curva verde vengono così approssimati malamente. Invece la seconda metà, a destra, è stata campionata con 4 bit e poiché $2^4 = 16$, le misure possono assumere 16 valori, approssimando così meglio la traccia. La figura è tratta dal rispettivo tutorial di audacity.

³Con la notazione scientifica i numeri si rappresentano in termini di potenze di 10, ad esempio il numero 3'140'000'000 si rappresenta come 3.14×10^9 , che poi sarebbe 3.14 miliardi

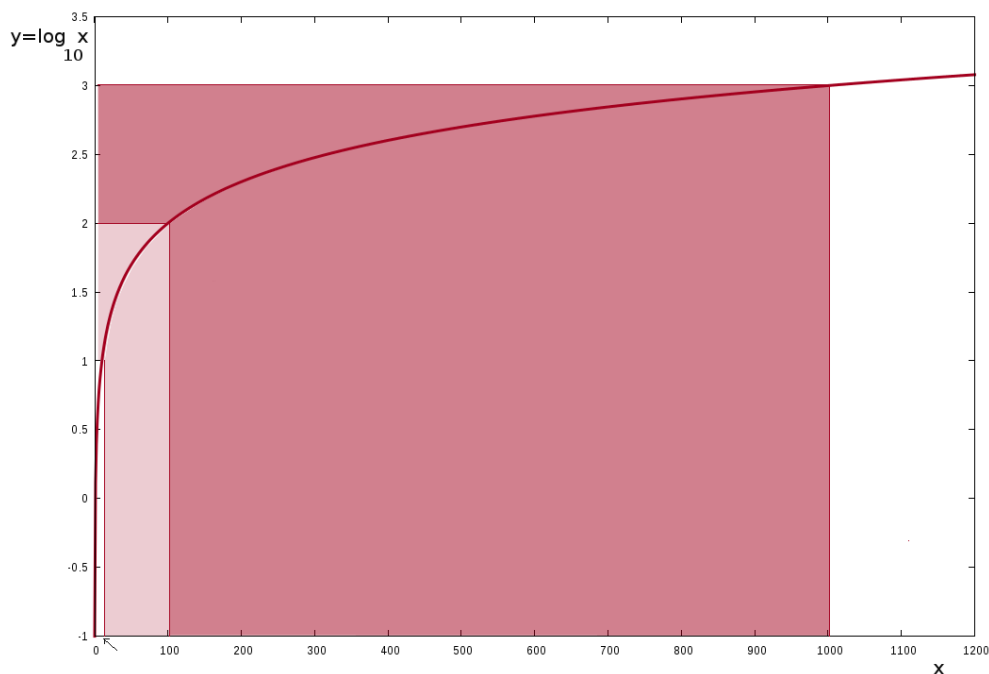


Figura 4: Grafico della funzione *logaritmo*.

verticale delle y). Con le funzioni matematiche si fanno viaggi sicuri: si sa di finire in luoghi diversi ma si sa come e dove e, sotto certe condizioni, si sa anche come tornare indietro. Tuttavia i mondi in cui si approda sono diversi e sovente sono proprio tali diversità che andiamo a cercare quando ricorriamo all'impiego di una funzione matematica. La figura 4 illustra come l'intervallo compreso fra 10 e 100 e quello compreso fra 100 e 1000 vengano trasformati nell'intervallo fra 1 e 2 e quello fra 2 e 3, rispettivamente. Qui si capisce come una quantità che subisce grandi variazioni, diciamo fra 10 e 1000, venga confinata in un ambito di variazioni molto più limitato dopo essere stata trasformata mediante la funzione logaritmo.

Abbiamo così i due ingredienti necessari per capire il *decibel* (qui ci riferiamo al caso dell'acustica, in realtà i decibel si impiegano in una varietà di altri contesti, laddove si abbia a che vedere con quantità che possono variare molto). Primo: esprimendo i valori di intensità mediante il rapporto con un valore di base di riferimento - soglia dell'udibile - si ottiene un numero puro che può essere utilizzato come argomento di una funzione matematica.

Secondo: applicando la funzione logaritmo a tale rapporto si ottiene l'effetto di ridurre la grande estensione delle possibili intensità. Chi è interessato può andare a vedere la definizione del decibel nella nota ⁴.

A noi che non siamo esperti di acustica il ricorso a questa unità di misura può parere inutile ma in realtà un fondamento c'è. Infatti la sensazione sonora che noi percepiamo è piuttosto di tipo logaritmico anziché lineare! Facciamo un gioco per rendercene conto. In questo audio ⁵ pronuncio la parola - audacity. L'audio è molto breve, della durata di un paio di secondi, tuttavia prima e dopo la parola c'è del silenzio che, a dire il vero, non è silenzio: se lo ascoltate con una buona cuffia (non con l'altoparlantino del portatile!) sentirete che invece del silenzio c'è un discreto fruscio, il cui volume non è così trascurabile rispetto al volume con il quale è stata pronunciata la parola.

Per sentirlo meglio può convenire caricare il file nel vostro Audacity e fare correre il brano ciclicamente - nel menu **Attività - Riproduci ciclicamente** oppure **Maiusc-spazio**. In Audacity c'è la possibilità di vedere le tracce sia in forma lineare che in decibel. Il menu dove possiamo specificare la visualizzazione delle tracce è quello che si apre cliccando sulla freccia accanto alla dicitura Traccia aud, nelle specifiche elencate a sinistra delle tracce, ma si può anche usare anche usare le scorciatoie da tastiera: **Maiusc-mb** per vedere le tracce in decibel, **Maiusc-mr** - facendo solamente **Maiusc-m** si apre il menu con le varie possibilità. Ebbene, nella figura 6 si confronta la forma d'onda in scala lineare con la versione in decibel, ovvero in scala logaritmica. Se guardate la figura ascoltando l'audio vi accorgete come la rappresentazione in decibel sia

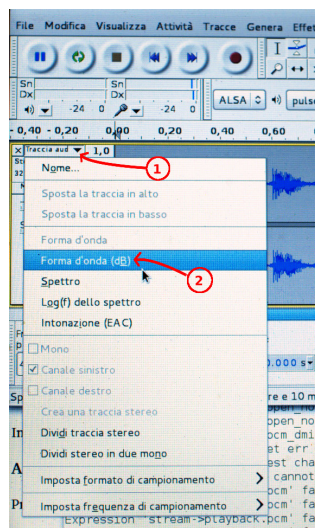


Figura 5: Voce di menu che consente di passare dal display lineare della traccia a quello espresso in decibel e viceversa.

⁴ La definizione esatta di decibel utilizzata nel campo dell'acustica è

$$dB = 20 \log \left(\frac{p}{p_0} \right) \quad (1)$$

dove \log è il logaritmo in base 10, p il valore medio della pressione sonora, più esattamente il valore quadratico medio, e p_0 la pressione sonora della soglia di udibilità, presa a riferimento.

⁵Chi sta leggendo il testo su carta può annotarsi l'URL per raggiungere successivamente l'audio: <https://iamarf.files.wordpress.com/2015/01/audacity.mp3>

più fedele.

Nella versione lineare (sopra) sembra che le zone di silenzio siano veramente tali, praticamente zero rispetto a quelle parlate, invece, la versione in decibel offre una rappresentazione molto più aderente a quella che è la nostra percezione sonora.

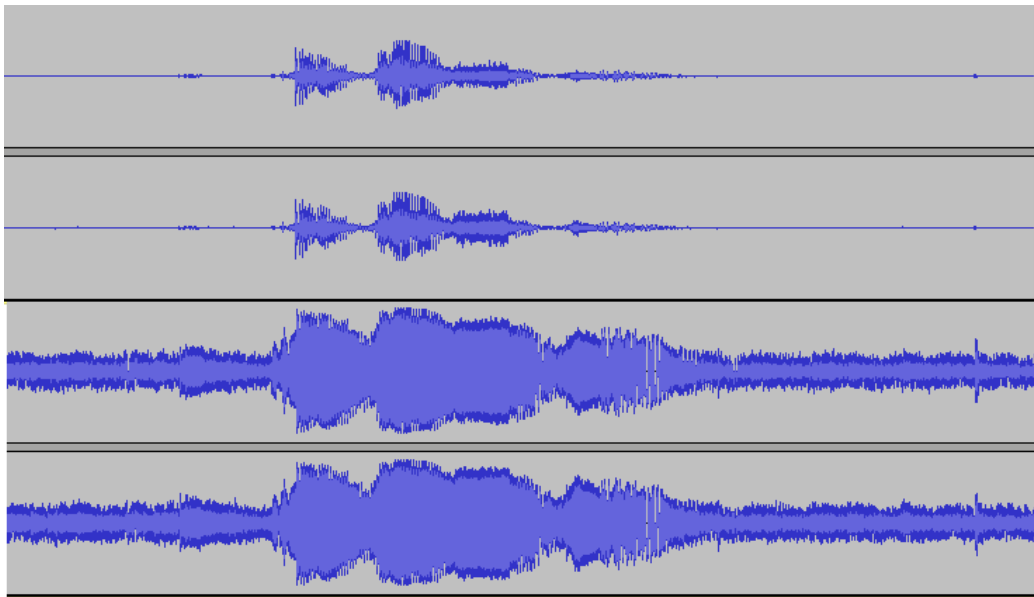


Figura 6: Le due tracce stereo della parola audacity. Sopra la forma d'onda (scala lineare), sotto in dB (scala logaritmica).

Questo è il motivo per cui nei software di trattamento dei segnali audio si utilizza volentieri il decibel. Consideriamo per esempio il controllo del guadagno del microfono. Prima di fare una registrazione occorre sempre controllare che il segnale non venga registrato troppo attenuato o troppo amplificato. Nel primo caso si può ovviare amplificandolo a posteriori per via digitale ma le intensità più deboli rischiano di venire amplificate in maniera distorta perché sono state quantizzate in maniera troppo discontinua. Nel secondo

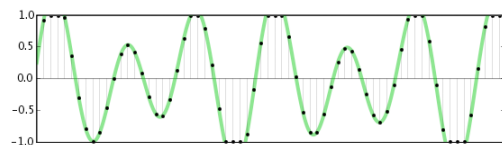


Figura 7: Illustrazione del fenomeno del clipping, che si ha quando a causa di un'amplificazione eccessiva la forma d'onda risulta tagliata orizzontalmente agli estremi inferiore e superiore. Un effetto difficile da correggere: meglio regolare preventivamente in maniera adeguata il guadagno del microfono.

caso, anche se è possibile attenuare il segnale per via digitale, l'operazione è destinata a fallire per via del fenomeno del *clipping*, che consiste nell'appiattimento della forma d'onda dove questa supera il massimo consentito dalla strumento.

In Fig. 8 sono evidenziati la finestra di controllo del segnale in ingresso e il comando per la regolazione del guadagno del microfono

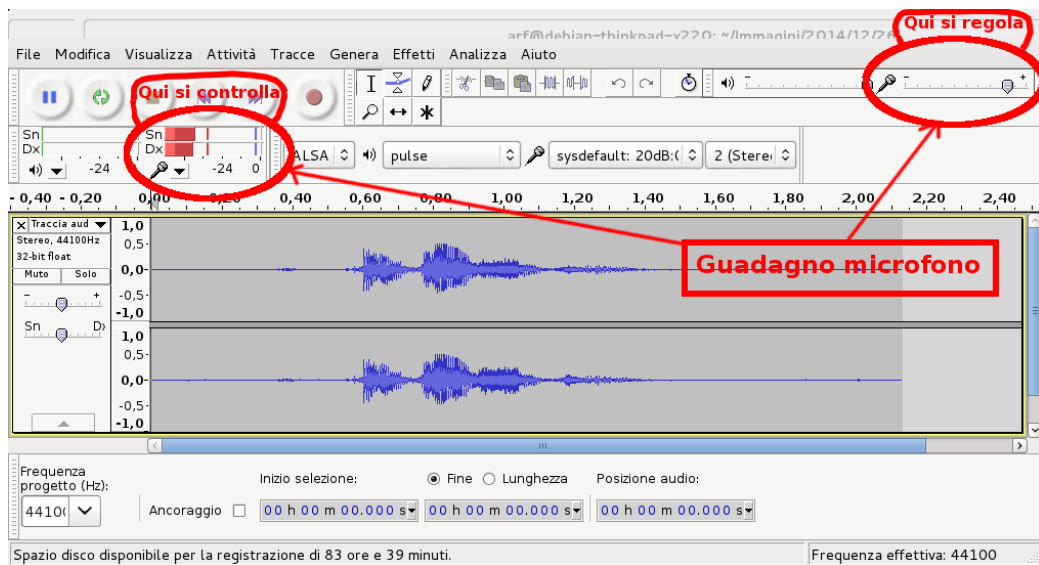


Figura 8: A sinistra è evidenziata la finestra di controllo del segnale in ingresso e a destra il comando per regolare il guadagno del microfono.

Per vedere in tempo reale l'ampiezza del segnale nella finestra di controllo basta cliccarci dentro. Le bande rosse che vedete in figura rappresentano grosso modo l'attuale livello di rumore nella mia stanza. L'estensione delle bande rappresenta il rapporto fra il massimo segnale possibile e l'attuale, quindi un numero minore di 1. Cosa significano quindi quel -24 e quello 0 sulla scala? Sono i valori espressi in decibel. Se chiamo D quel valore allora (vedi nota 4) $D = 20 \log \left(\frac{I}{I_m} \right)$, dove I è l'intensità del segnale e I_m il suo massimo valore possibile. Per vedere quanto vale il rapporto I/I_m si deve invertire l'equazione e sostituire in luogo di D il valore -24 a: $10^{-24/20} = 0.063$. Ancora: se invece fosse $D = -20$ allora avremmo $10^{-20/20} = 10^{-1} = 1/10 = 0.1$. Infine possiamo verificare che quando $D = 0$ allora $10^{0/20} = 10^0 = 1$. In sintesi a parole: la finestra di controllo informa sul valore dell'intensità del suono rispetto al massimo consentito dallo strumento. Quando il valore è uguale al massimo il rapporto vale 1 e la corrispondente espressione in decibel vale

0. Se per esempio il rapporto è un decimo, 0.1, il valore in decibel è -20 .
Eccetera.

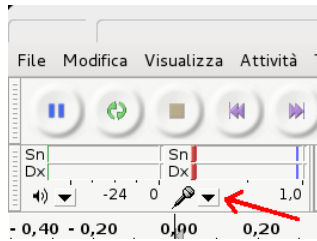


Figura 9: Come appare il controllo dell'input una volta selezionata la scala lineare; la freccia indica dove si apre il menu che consente di gestire il controllo, fra cui la scelta fra scala lineare e scala in decibel

Cliccando sull'icona indicata in figura 9 si può passare dalla scala lineare a quella in decibel e viceversa. La figura rappresenta il caso della scala lineare con lo stesso livello di rumore ambientale della figura 8. Si vede come la larghezza delle bande risulti molto più piccola e come, ovviamente, la scala termini al valore 1.0.

Livelli indicativi di vari tipi di rumore espressi in decibel	
dB	Sorgente
300	Eruzione del Krakatoa nel 1883
250	All'interno di un Tornado
180	Razzo al decollo
140	Colpo di pistola a 1 m, auto di Formula 1
130	Soglia del dolore
125	Aereo al decollo a 50 m
120	Sirena
110	Motosega
100	Discoteca, concerto rock
90	Urlo, fischiello
80	Camion pesante a 1 m
70	Aspirapolvere a 1 m; radio ad alto volume
60	Ufficio rumoroso, radio, conversazione
50	Ambiente domestico; teatro a 10 m
40	Quartiere abitato, di notte
30	Sussurri a 1 m
20	Respiro umano
0	Soglia dell'udibile
-9	Camera anecoica

Tabella 1: Trascrivendo questa tabella mi rendo conto come il *tinnito* che mi fa sempre compagnia abbia un'intensità intorno a 20 *Db*. Ne informerò l'otorino.