

UNIVERSITA' DEGLI STUDI DI MILANO

FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E
NATURALI



CORSO DI LAUREA IN SCIENZE E TECNOLOGIE

DELLA COMUNICAZIONE MUSICALE

METODI PER LA VALUTAZIONE
DELL'INTENSITA' PERCEPITA NEI SEGNALI
AUDIO

Relatore:
Prof. Luca A. Ludovico

Correlatore:
Dott. Davide A. Mauro

Laureando:
Niklas D'Antoni

Matricola 709097

Anno Accademico 2008/2009

➤ **INTRODUZIONE.....4**

➤ **CAPITOLO 1: SUONO E PERCEZIONE**

1.1)	CONCETTI FONDAMENTALI	7
1.2)	INTRODUZIONE ALLA PSICOACUSTICA	10
1.3)	STRUTTURA DELL'ORECCHIO	11
1.4)	SENSAZIONE SONORA	18
	<i>1.4.1) Intensità sonora percepita: il phon</i>	19
	<i>1.4.2) Dal phon al son</i>	21
	<i>1.4.3) Altezza in funzione della frequenza: il mel</i>	22
	<i>1.4.4) Bande critiche</i>	23
	<i>1.4.5) Il mascheramento</i>	24

➤ **CAPITOLO 2: LOUDNESS: DAI PRIMI STUDI FINO ALLO
STANDARD ATTUALE OSI 226**

2.1)	LAVORO DI HARVEY FLETCHER E W. A. MUNSON	27
	<i>2.1.1) Definizioni</i>	28
	<i>2.1.2) Metodo sperimentale per la misurazione del livello di loudness di un suono statico</i>	30

2.1.3) <i>Confronto con il lavoro di B.A. Kingsbury</i>	35
2.2) GLI STUDI DI D. W. ROBINSON E R. S. DADSON	38
2.3) STANDARD ISO 226	40
2.3.1) <i>Revisione dello standard: le nuove curve di Yôiti Suzuki</i>	42

➤ **CAPITOLO 3: INTENSITA' SONORA PERCEPITA NEI
SUONI COMPLESSI**

3.1) CALCOLO DELLA LOUDNESS PER TONI COMPLESSI.....	49
3.2) CONTROLLO DELLA LOUDNESS.....	52
3.2.1) <i>Loudness nei programmi televisivi</i>	52
3.2.2) <i>The Loudness War</i>	53

➤ **CONCLUSIONI**56

➤ **BIBLIOGRAFIA**58

INTRODUZIONE

Il presente elaborato tratta un argomento che negli ultimi decenni è tornato ad interessare ricercatori di tutto il mondo. Lo studio per la ricerca di metodi e tecnologie in grado di valutare l'intensità sonora percepita nei segnali audio, ha iniziato a catturare l'attenzione di numerosi scienziati già agli inizi dello scorso secolo; il progresso tecnologico aveva finalmente dato la possibilità di spiegare come il nostro orecchio e la nostra mente reagiscano agli stimoli sonori. Dopo un forte interesse iniziale, che vide studiosi come Fletcher e Munson gettare le basi per i futuri studi, l'argomento venne accantonato per alcuni decenni. Fu poi negli anni '80 e '90 che la questione venne riesaminata; le nuove e più potenti tecnologie ora disponibili assieme ai progressi fatti nell'ambito della psicoacustica nel corso degli anni, hanno portato numerosi scienziati a rivedere i lavori precedentemente svolti ed ad iniziarne di nuovi con l'obiettivo di ottenere risultati più affidabili e precisi.

Lo scopo di questa tesi è quello analizzare i diversi progetti di ricerca, in merito all'intensità percepita nei segnali sonori (loudness), che si sono succeduti negli anni dal secondo decennio del ventesimo secolo fino ai giorni nostri. L'oggetto centrale di questa ricerca sono le così dette curve di uguale intensità sonora percepita, meglio conosciute come *curve isofoniche*. Queste curve tracciano un grafico di come l'apparato uditivo umano percepisce in maniera differente l'intensità di un suono a seconda della frequenza che lo caratterizza. Per poter meglio comprendere come ciò possa avvenire, si è scelto di dedicare la prima parte dell'elaborato all'introduzione di quei concetti di acustica e psicoacustica che stanno alla base di queste ricerche.

Come prima cosa viene brevemente introdotto il concetto di suono e di tutti quegli elementi che ne caratterizzano la forma d'onda. Ben più rilievo è stato dato agli elementi di psicoacustica ed in particolar modo alla struttura del nostro apparato uditivo ed alla sua affascinante complessità; il funzionamento combinato di ogni suo singolo componente ne fa oggetto di studio da centinaia di anni. Più in particolare, il fatto che l'orecchio risponda

in maniera non lineare agli stimoli acustici che lo raggiungono, gli conferisce un ruolo centrale nel giudizio dell'intensità sonora percepita. Conoscere i principi fisici che regolano il funzionamento dell'orecchio, è essenziale per comprendere come esso reagisce agli stimoli sonori e gli effetti che ne derivano.

Prima di passare alla seconda parte dell'elaborato, che prende in esame gli studi fatti per l'elaborazione delle curve isofoniche, si è voluto introdurre altri concetti di natura fisica e psicofisica che riguardano parametri di riferimento per la misurazione di diverse caratteristiche del suono; unità di misura come il *Decibel*, per misurare la pressione sonora, gli *Hertz*, per la misurazione dell'altezza, e i *Phon*, per misurare l'intensità percepita dei suoni, vanno ben compresi prima di affrontare uno studio sulle curve isofoniche. Vengono inoltre introdotti il concetto di *banda critica* ed il fenomeno del *mascheramento* che, come si vedrà nel terzo capitolo, andranno tenuti in considerazione nel calcolo della loudness dei suoni complessi.

Una volta chiariti i concetti fondamentali, la tesi si concentra sugli studi che a partire dal 1924 hanno interessato l'argomento. Il capitolo inizia facendo una panoramica dei primi studi compiuti sulla loudness per poi soffermarsi sugli esperimenti compiuti da Fletcher e Munson nel 1933. Il lavoro di questi due ricercatori ha un'importanza storica: furono i primi a stilare delle curve isofoniche ottenute da test audiometrici svolti su un campione di persone normoudenti.

Il lavoro di Fletcher e Munson venne poi ripreso, nel 1956, da Robinson e Dadson che, sulla base di nuovi esperimenti, crearono un nuovo set di curve isofoniche. Queste curve vennero considerate per molti anni le più affidabili ed andarono a costituire lo standard internazionale OSI 226. Tuttavia, nuove ricerche effettuate negli anni '80 e '90, evidenziarono gravi difetti in queste curve e spinsero ricercatori come Yôiti Suzuki a riaffrontare l'argomento. L'obiettivo di Suzuki era quello di creare un nuovo e più accurato set di curve che andasse a modificare lo standard OSI 226 attuale, e ci riuscì dimostrando, tra l'altro, che le nuove curve isofoniche da lui create sulla base di dati provenienti da varie ricerche, hanno più tratti in comune con le curve di Fletcher e Munson

(1933) che con quelle di Robinson e Dadson (1956) che hanno costituito lo standard OSI 226 per diverso tempo.

L'ultimo capitolo tratta come argomento principale l'intensità sonora percepita nei segnali complessi. Il calcolo della loudness per un suono complesso è ben più complicato rispetto ad un tono puro perché entrano in gioco fenomeni come il *mascheramento*, già descritto nel primo capitolo. Tuttavia, esistono due metodi, che verranno poi descritti, per il calcolo dell'intensità sonora percepita nei suoni complessi. I due metodi, previsti dallo standard OSI 532, sono il metodo di Stevens (più semplice e diffuso) ed il metodo di Zwicker.

L'elaborato si conclude analizzando come la loudness viene oggi gestita in ambito musicale e televisivo; in particolare si farà riferimento al fenomeno della *loudness war* che è diventata oggi forte oggetto di discussione specialmente in ambito discografico.

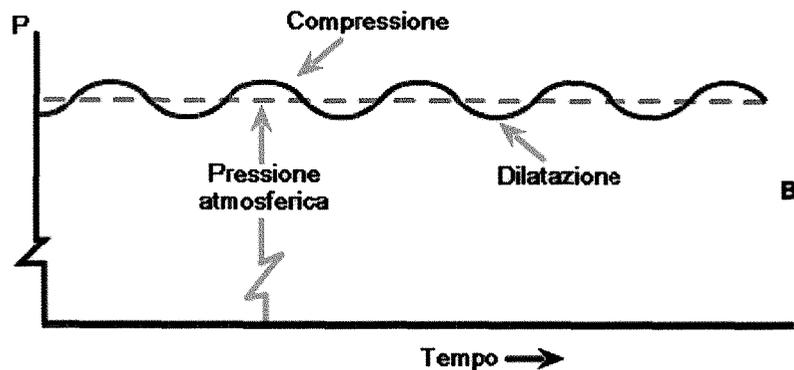
CAPITOLO 1

SUONO E PERCEZIONE

1.1) CONCETTI FONDAMENTALI

Con il termine "suono" si tende generalmente ad indicare la combinazione di due fenomeni: la propagazione in un mezzo elastico di onde prodotte da una sorgente messa in vibrazione e l'interpretazione di queste onde, da parte dell'uomo, mediante il complesso sistema orecchio-cervello. Affinché un suono si possa propagare ha bisogno di un mezzo che lo trasporti: qualsiasi mezzo solido, liquido o gassoso che abbia proprietà di elasticità nel movimento delle sue particelle elementari. Fondamentalmente, ciò che caratterizza un suono da un punto di vista strettamente fisico, è la frequenza che si misura in cicli al secondo, ovvero *Hertz (Hz)*, legata alla velocità di vibrazione della sorgente e l'intensità, legata all'ampiezza o più propriamente all'energia dell'oscillazione stessa, che si misura in *deciBel (dB)*.

Mentre l'orecchio esterno rileva la variazione periodica di pressione, l'orecchio interno la converte in impulsi elettrici da inviare al cervello che interpreta tale variazione come suono. La propagazione del suono è affidata allo spostamento della massa d'aria che circonda la sorgente sonora che lo produce, ciò significa che in assenza di un mezzo elastico in grado di trasportare l'onda generata dalla sorgente sonora (come ad esempio nello spazio), il suono non esiste. Il fluido che circonda la sorgente sonora (in genere l'aria) ha una pressione costante, cioè in ogni unità di volume vi è lo stesso numero di molecole. Quando l'aria viene perturbata, il valore di pressione varia da punto a punto: aumenta dove le molecole sono compresse, diminuisce dove le molecole sono espanse.

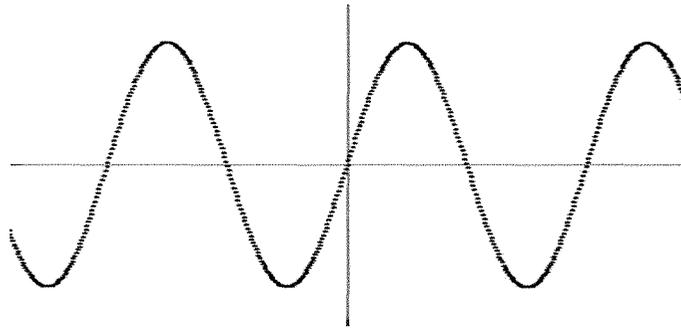


Il nostro orecchio percepisce la presenza di una emissione sonora proprio rilevando la variazione di pressione ambientale. Per questo, tale variazione prende il nome di *pressione sonora*. L'unità di misura della pressione sonora assoluta è il *pascal* (simbolo Pa, equivalente a 1 newton per metro quadro) ma, in considerazione delle caratteristiche di percezione dell'apparato uditivo di un essere umano, è spesso consigliabile ricorrere ad una descrizione in termini relativi, cioè posta in relazione con un valore di riferimento. In generale, come riferimento viene adottato quella che si usa definire *la soglia di percezione uditiva*, numericamente pari a $20 \mu\text{Pa}$, ossia 20 milionesimi di pascal. L'unità di misura della pressione sonora relativa è il *decibel* (simbolo *dB* spesso accompagnato dal suffisso *spl*, acronimo del termine inglese *sound pressure level* ossia livello di pressione sonora). Questa unità di misura deve il suo nome allo scienziato statunitense che per primo ne fece uso, Graham Bell. Inizialmente l'intensità sonora era misurato in *Bell*, ci si accorse presto però che l'orecchio umano percepisce la pressione sonora in maniera logaritmica anziché lineare, quindi risulta conveniente esprimere le grandezze legate all'ampiezza del suono in un'unità di misura logaritmica: il *decibel*.

Un altro concetto che va introdotto è quello di *soglia differenziale*. Il concetto venne introdotto da Weber nel 1834 ed indica la differenza minima di intensità che deve esistere tra due stimoli perché possano essere distinti. Weber stabilì una legge secondo cui la soglia differenziale per l'intensità di uno stimolo, è una proporzione costante dell'intensità dello stimolo di riferimento; indipendentemente dalla grandezza di due stimoli, il secondo deve differire dal primo di una porzione costante affinché possa essere percepito come differente. Per quanto riguarda l'orecchio, un cambiamento di intensità di 1 *dB* non viene

percepito, sono necessari almeno 3 *dB* per la percezione. Con 5 *dB* si ha una percezione chiara del cambiamento avvenuto mentre per il raddoppio o per il dimezzamento dello stimolo ricevuto sono necessari circa 10 *dB*.

Come già detto in precedenza, la sollecitazione della sorgente sonora crea un'onda nel fluido circostante. Le forme d'onda possono arrivare ad essere molto complicate, ma tutte possono essere considerate un'estensione di una forma d'onda molto semplice: *la sinusoidale*.



Il *periodo* è la distanza tra due creste ed equivale al tempo impiegato dalla particella per tornare nello stesso punto dopo aver cominciato lo spostamento (indica cioè la durata di un'oscillazione completa). La distanza dalla cresta all'asse delle ascisse indica, invece, l'*ampiezza* del movimento, in altre parole la distanza massima percorsa dalla particella dalla sua posizione di riposo durante l'oscillazione. Tuttavia, nonostante il periodo e l'ampiezza siano due grandezze che da sole sarebbero sufficienti per descrivere le caratteristiche di un'onda, non sono frequentemente utilizzate, perlomeno non in forma pura: in acustica si preferisce infatti usare altre grandezze da queste derivate. Dal numero di periodi compiuti in un secondo si ottiene la frequenza, misurata in Hertz (*Hz*), che è definita proprio come il numero di oscillazioni compiute dalla particella in un secondo. Dall'ampiezza dell'onda, invece, si calcola la pressione sonora, definita come la variazione di pressione rispetto alla condizione di quiete.

1.2) INTRODUZIONE ALLA PSICOACUSTICA

Studiando il sistema uditivo umano è facile dedurre che la percezione delle stimolazioni acustiche dipende da complicati meccanismi fisiologici e psicologici. Mentre i meccanismi per i quali le vibrazioni acustiche dell'ambiente aereo esterno vengono tradotte in impulsi nervosi sono discretamente conosciuti, così come quelli della conduzione di questi dai recettori periferici al cervello, non altrettanto si può dire dei meccanismi che trasformano gli impulsi in sensazioni coscienti. Gli aspetti psicologici sono studiati dalla psicoacustica, il cui compito è quello di stabilire le relazioni intercorrenti tra grandezze acustiche “oggettive” e le valutazioni “soggettive”, ossia la percezione, del suono. La psicoacustica è una branca della psicofisica che tenta quindi di quantificare gli aspetti psicologici dei suoni. Essa si occupa della descrizione dei limiti e delle modalità attraverso cui si realizza il senso dell'udito in relazione alle caratteristiche fisiche dei segnali acustici, in particolare alle differenze dei valori dei parametri acustici percepibili dall'orecchio umano. Per questi fini si fa ricorso al concetto di soglia. Per valore di soglia assoluta di un segnale sonoro si intende il valore oltre al quale si ha una risposta da parte del sistema uditivo e al di sotto della quale non si ha alcuna risposta. La soglia non è un parametro costante per tutti gli individui e dipende, oltre che dalle condizioni fisiche e mentali dell'ascoltatore, anche dalle caratteristiche acustiche dello stimolo sonoro quali la frequenza, l'intensità e la durata. Il concetto di soglia è molto importante e sta alla base degli studi che verranno analizzati nel secondo capitolo.

Per chiarire i concetti di psicoacustica e nello specifico della loudness, il seguente capitolo è dedicato alla fisiologia dell'apparato uditivo dando risalto maggiore al ruolo dell'orecchio interno nel quale avviene la trasduzione vibratoria del suono in impulsi elettroneurali che ci permettono di percepire la sensazione sonora.

1.3) STRUTTURA DELL'ORECCHIO

Con il termine orecchio si prende in considerazione l'intero apparato uditivo; dalla parte visibile ai lati della nostra testa (il *padiglione auricolare*) fino alle complesse strutture dell'orecchio interno, collocate profondamente nel cranio.

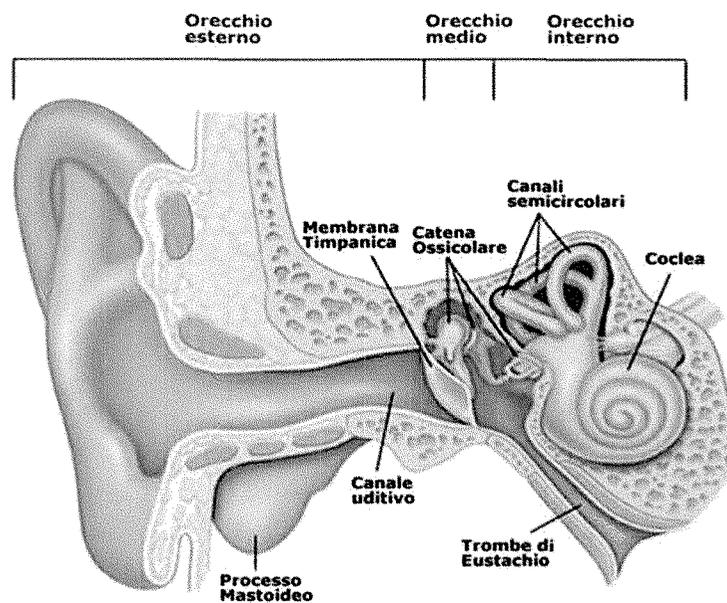
Il sistema uditivo viene solitamente diviso in tre distinte strutture:

- *l'orecchio esterno*, costituito dal padiglione auricolare e dal canale uditivo, svolge il compito di convogliare il suono verso l'orecchio medio. Il padiglione auricolare raccoglie il suono e fornisce al cervello elementi per la valutazione della direzionalità del suono. Esso cioè sovrappone al suono stesso informazioni (sotto forma di frequenze) riguardo alla direzione dalla quale il suono è giunto al padiglione stesso. Ciò spiega la forma particolarissima del padiglione auricolare umano, che si distingue da quello degli altri animali che spesso è direzionabile. Si può condurre un esperimento psicoacustico per dimostrare questa capacità: si ascolta con un auricolare (saltando cioè il filtro del padiglione, e perciò ingannando l'orecchio) la gamma di un'ottava di rumore casuale. Se regoliamo la frequenza centrale a $7,2\text{ kHz}$, il rumore sembra provenire da una sorgente all'altezza dell'osservatore. Se la regoliamo a 8 kHz , il suono sembra provenire da una sorgente posta in alto. Se invece regoliamo la frequenza centrale a $6,3\text{ kHz}$, il suono sembra provenire dal basso. Il canale uditivo, invece, conduce le informazioni sonore raccolte dal padiglione alla membrana del timpano. Le dimensioni del canale uditivo indicano che esso ha una frequenza di risonanza a 3000 Hz (lunghezza d'onda 12 cm , 4 volte la lunghezza del canale) in corrispondenza della quale si ha un picco nella risposta in frequenza dell'orecchio. Un picco secondario si ha a 9000 Hz . Il picco in risposta ai 3000 Hz indica che l'orecchio è specificamente "tarato" sulla voce umana.

- *l'orecchio medio*, che comprende il timpano seguito dalla catena degli ossicini: martello, incudine e staffa ha il compito di trasmettere lo stimolo sonoro all'orecchio interno. In questa parte risiede anche la tromba di Eustachio che è il canale di collegamento tra la

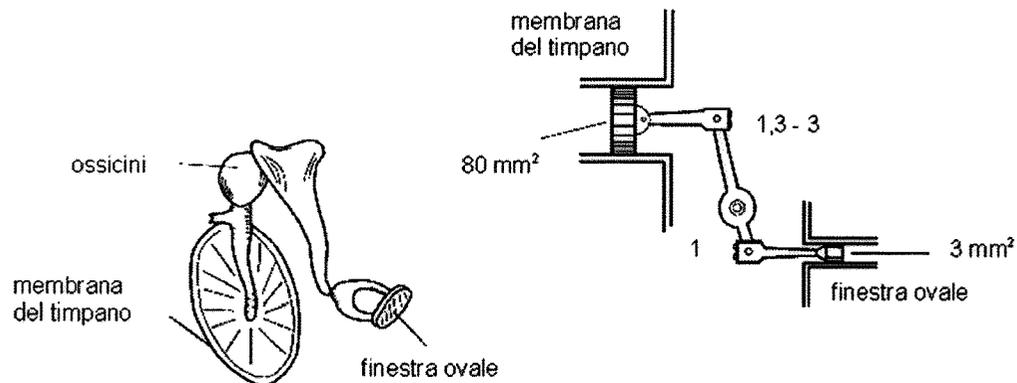
cassa timpanica e la faringe ed ha la funzione di mantenere il corretto equilibrio di pressione tra orecchio esterno e medio;

- *l'orecchio interno*, collocato all'interno dell'osso temporale, è costituito dall'apparato vestibolare, responsabile dell'orientamento, e dalla coclea, nella quale avviene la trasduzione della vibrazione meccanica in segnale elettrico, che giunge al cervello tramite il nervo acustico.



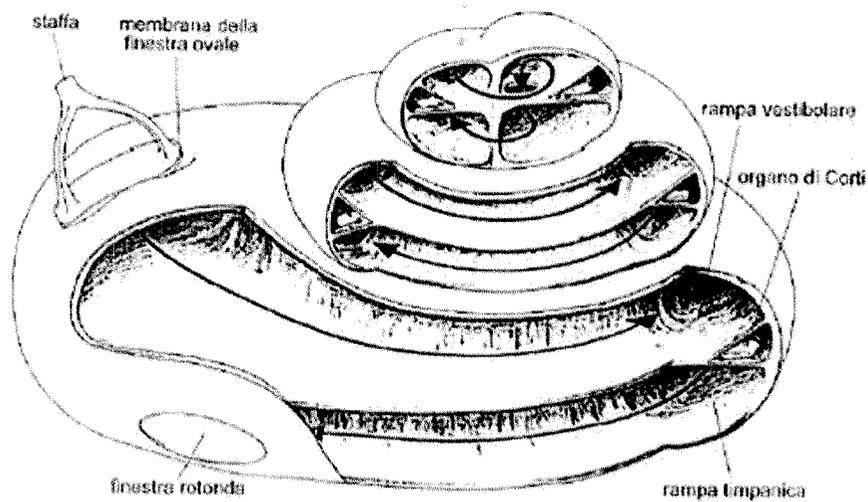
Il comportamento dell'orecchio umano, sottoposto a stimoli sonori, si può schematizzare come segue: l'orecchio esterno capta gli stimoli sonori e li convoglia all'interno del canale uditivo sino all'orecchio medio. A questo punto avviene la trasmissione delle variazioni di pressione che vengono trasformate in vibrazioni meccaniche del timpano e, tramite la catena di ossicini, del liquido cocleare; l'orecchio interno, con i raffinati meccanismi presenti nella coclea, effettua la trasduzione di tali vibrazioni in potenziali d'azione che trasportano così l'informazione contenuta nel suono al cervello. In questo processo è da notare però che, essendo l'impedenza acustica del liquido molto maggiore di quella dell'aria si rende necessario un adattamento d'impedenza tra i due mezzi, così da minimizzare la riflessione della vibrazione nel suo passaggio all'orecchio interno. La

grande superficie del timpano rispetto alla finestra ovale (dell'ordine di 15:1) e il rapporto di riduzione della catena di ossicini assolvono proprio a questa funzione di adattamento, riducendo l'ampiezza dei movimenti in modo da rendere la pressione sul fluido cocleare maggiore di quella presente sulla membrana timpanica.

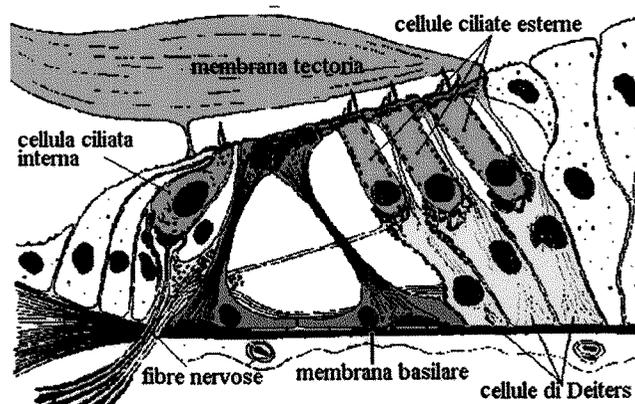


La coclea è composta essenzialmente da una spirale ossea avvolta per due giri e mezzo attorno ad un perno centrale detto *modiolo*. E' suddivisa in tre *scale* o *rampe*: *scala vestibolare*, *scala media* (o dotto cocleare) e *scala timpanica*, separate da due membrane (membrana basilare e membrana di Reissner). La scala vestibolare e la scala timpanica sono messe in comunicazione dall'*elicotrema*, un foro presente nella zona apicale della coclea, che permette il passaggio della perilinfia, il fluido ricco di ioni sodio in esse contenuto. La scala media termina invece prima dell'apice della coclea e contiene un liquido, l'endolinfia, simile a quello delle scale esterne, ma con alta concentrazione di ioni potassio. La differente composizione chimica dei due fluidi fa sì che tra questi esista una differenza di potenziale di circa 90mV, che permette di alimentare i meccanismi di amplificazione elettromeccanica presenti all'interno della coclea. La vibrazione acustica viene trasferita all'orecchio interno dalla staffa che tramite la finestra ovale produce onde di compressione e rarefazione della perilinfia contenuta nel canale cocleare. Tali oscillazioni, essendo la membrana di Reissner estremamente sottile e flessibile, vengono prontamente trasmesse alla scala media e, da questa, alla membrana basilare. Durante l'onda di compressione la membrana basilare tenderà quindi a flettersi verso la scala timpanica mentre durante l'onda di rarefazione tenderà a flettersi verso la scala vestibolare.

Le caratteristiche visco-elastiche della membrana basilare, fanno sì che le frequenze più elevate (suoni acuti) facciano oscillare più efficacemente quelle porzioni di membrana poste vicino alla finestra ovale mentre le frequenze più basse (suoni gravi) facciano oscillare le porzioni terminali della membrana stessa, vale a dire le porzioni poste verso le regioni apicali della coclea.



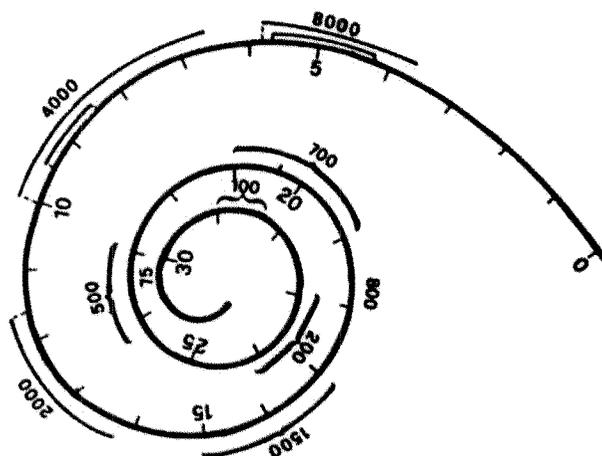
Sulla membrana basilare è posta la parte sensoriale dell'orecchio, l'organo di Corti. Esso è costituito sia da cellule di sostegno sia da cellule sensoriali. Le cellule sensoriali dell'organo di Corti sono dette *cellule cigliate*, mentre quelle di sostegno *stereociglia*.



E' proprio questa la parte più complessa e delicata dell'apparato uditivo. Essa analizza i suoni, separando e facendone risaltare le componenti significative, smorzandone le componenti più intense e accentuandone quelle più deboli, favorendo la rivelazione dei dettagli acustici che permettono di individuare le caratteristiche delle sorgenti sonore. La parte sensoriale (vale a dire quella che analizza le onde sonore) della coclea è rappresentata, appunto, dall'organo di Corti che è poggiato sulla membrana basilare e si estende per tutta la lunghezza della coclea stessa. L'organo di Corti è costituito da numerose popolazioni cellulari di cui, da un punto di vista funzionale, le più importanti sono le cellule sensoriali (cellule cigliate). L'attivazione delle cellule cigliate è, come tutti i processi biologici, affidata ad una serie di meccanismi ionici e neuro-secretori assai complessi.

Il punto in cui la deformazione della membrana raggiunge la massima ampiezza di picco, che dipende dalla frequenza, corrisponde ad una regione del sovrastante organo di Corti che è maggiormente eccitata, in tal modo, solo un piccolo gruppo di cellule cigliate sono sollecitate da quella particolare frequenza sonora. In corrispondenza di questo picco si creano dei vortici nella perilinfa delle due rampe del canale cocleare. La sua ampiezza declina poi rapidamente proseguendo verso l'apice della coclea. A causa delle variazioni continue delle sue proprietà meccaniche, la membrana è sintonizzata in ogni suo punto per una diversa frequenza.

Le frequenze caratteristiche variano in modo monotono e continuo in funzione della posizione lungo la coclea, da un minimo di 20 Hz all'apice ad un massimo di 20 kHz alla base. Tale relazione non è lineare ma logaritmica e questa localizzazione puntuale delle frequenze sonore lungo la membrana, costituisce la mappa tonotopica. In definitiva, mentre uno strumento musicale sintetizza un suono complesso partendo dalla combinazione di suoni puri, l'orecchio opera all'inverso, analizzando i suoni complessi nelle loro componenti pure.



Assai interessante è il ruolo giocato dalle cellule cigliate interne ed esterne nella funzione uditiva. Le cellule cigliate esterne percepiscono le vibrazioni della membrana basilare mediante i loro fasci di stereociglia. Le stereociglia si trovano immerse nell'endolinfa, un fluido che possiede un potenziale elettrico di +80 mV rispetto a quello della perilinfia che circonda la parte inferiore della cellula. Sempre rispetto a questo, l'interno delle cellule possiede invece un potenziale negativo di -70 mV, mantenuto dalle potenti batterie a sodio e potassio delle membrane cellulari. Pertanto, attraverso le stereociglia si forma una differenza di potenziale di circa 150 mV capace di generare correnti elettriche fino a 8 nA che causano la contrazione della cellula.

Le cellule cigliate interne sono localizzate nell'organo di Corti verso l'interno rispetto all'asse della chiocciola, queste cellule (non contrattili) sono i sensori che trasmettono i segnali al nervo acustico. Da questo si deduce come le cellule cigliate interne siano il vero e proprio trasduttore meccanico elettrico. Il ruolo delle cellule cigliate esterne è quello di supporto della membrana basilare amplificando le piccole vibrazioni e smorzando quelle intense.

La membrana basale funge inoltre da filtro molto selettivo per quanto riguarda le frequenze percettibili; essa è infatti molto tesa e sottile nelle vicinanze della finestra ovale, mentre diventa sempre più spessa e molle all'avvicinarsi del centro. La tensione della membrana è proporzionale alla frequenza di risonanza a cui lavora: i suoni ad alta frequenza vengono

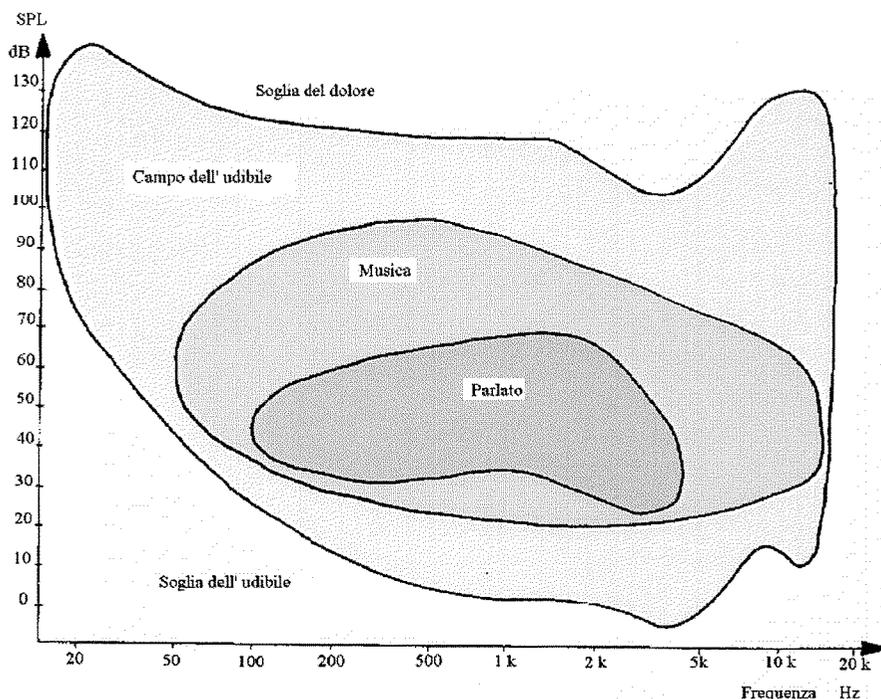
quindi riconosciuti subito, mentre quelli a bassa frequenza devono percorrere tutto il canale vestibolare (ca. 30 mm) prima di essere uditi. Questo fenomeno provoca l'attenuazione dei suoni a basse frequenze e spiega l'andamento delle curve isofoniche del diagramma di Fletcher e Munson che verranno analizzate nel dettaglio più avanti.

Un altro aspetto importante che si può osservare nell'apparato uditivo umano risiede nel fatto che il canale di trasmissione del suono è unico e assai limitato: questo fa sì che, quando sono presenti al suo interno troppe informazioni, quelle aventi minore intensità vengono trascurate. Questo accade quando un suono è caratterizzato da diverse componenti, simili in frequenza, ma in cui una prevalga per intensità: la sensazione che riceviamo coincide quindi con la componente più intensa, mentre le altre non sono percepibili. Il fenomeno appena descritto, che vedremo meglio nel prossimo capitolo, è detto *mascheramento*, ed è anche alla base delle più moderne tecnologie di compressione dei file audio.

1.4) SENSAZIONE SONORA

Come annunciato, la risposta dell'orecchio umano ad uno stimolo sonoro è in generale un fenomeno molto soggettivo.

Sottoponendo individui normoudenti a diversi stimoli sonori, dei quali vengono variate pressione sonora e frequenza, e analizzando le risposte uditive, è possibile tracciare un grafico, detto diagramma di sensazione sonora, che mostri quali suoni siano percettibili dall'uomo. Il limite inferiore del grafico (*soglia di udibilità*), rappresenta le pressioni minime, alle diverse frequenze, che vengono percepite dall'uomo; si può notare come suoni che abbiano un frequenza tra i 2 ed i 5 kHz necessitino di un'ampiezza minore per essere uditi, l'argomento verrà approfondito nei prossimi capitoli. Il limite superiore (*soglia del dolore*), indica invece la massima intensità sonora che può essere sopportata dall'orecchio senza innescare la sensazione del dolore. Fra queste due linee si estende l'area della sensazione uditiva, che comprende tutti suoni udibili.



Dal punto di vista spettrale, l'intervallo di frequenze udibili varia tra 10-20 Hz e 16-20 kHz. Le frequenze al di sotto della più bassa udibile si dicono *infrasuoni*, quelle al di sopra della più alta si dicono *ultrasuoni*.

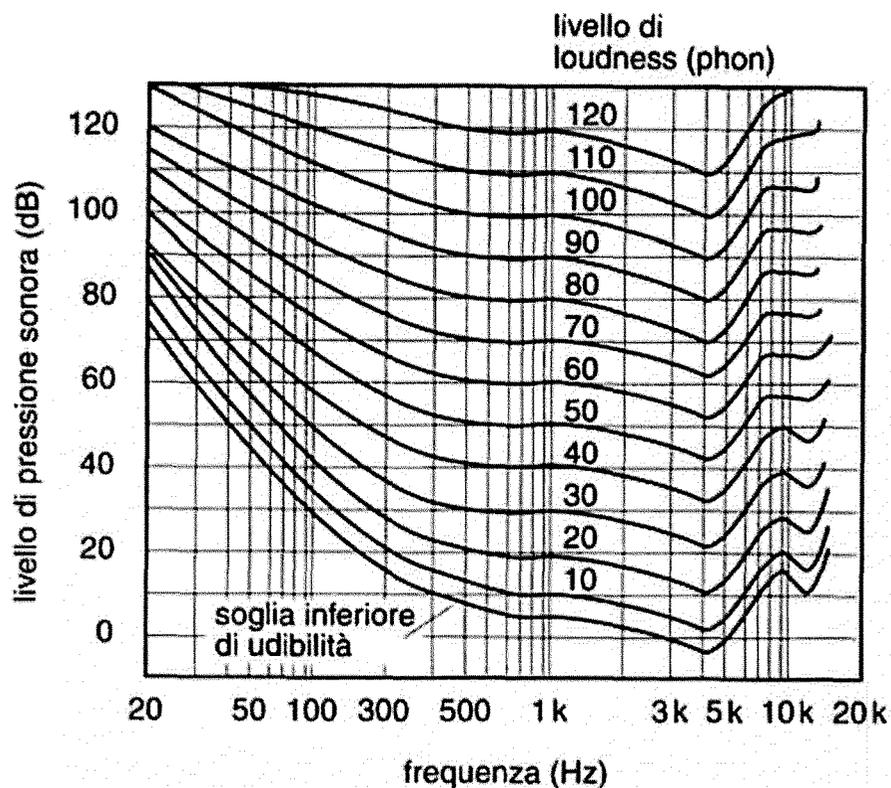
In questo intervallo (quello udibile) la sensibilità agli stimoli sonori è diversa: è minore alle basse frequenze, presenta un massimo verso i 4 kHz per poi tornare ad abbassarsi in maniera più lieve a frequenze maggiori; l'orecchio è quindi maggiormente sensibile alle frequenze elevate. La risposta dell'orecchio agli stimoli di diversa frequenza, varia fortemente negli anni, in particolare si nota una netta diminuzione della sensibilità alle alte frequenze (oltre i 10-12 kHz) con l'aumentare dell'età.

1.4.1) Intensità sonora percepita: il phon

Di fondamentale interesse per questo elaborato, è lo studio dell'intensità sonora percepita che verrà ora brevemente introdotto per poi essere approfondito nel secondo capitolo.

Come vedremo bene nel dettaglio, l'unità di misura per l'intensità percepita (il *phon*) deriva da uno studio svolto da due ricercatori della Bell Company (*Harvey Fletcher E W. A. Munson*) nel 1933.

Lo studio venne affrontato con una metodologia che è quella tipica della psicoacustica: numerose persone furono sottoposte a fenomeni sonori, e ne furono raccolte le dichiarazioni riguardo alle loro sensazioni. In questo caso furono verificate le sensazioni riguardo all'ampiezza dei suoni (quello che viene comunemente detto "il volume") rispetto alla frequenza del suono. Le curve del grafico seguente sono infatti anche dette *curve di isofonia* ad intendere che ogni curva rappresenta un certo "volume" percepito. La tabella nata da questi esperimenti, e poi, che come vedremo, ulteriormente analizzata e modificata nel corso degli anni, costituisce oggi lo standard internazionale ISO 226: ciò è indicativo di quanto lo studio della psicoacustica non sia solo una curiosità psicologica, ma uno studio di importanza fondamentale per numerosi aspetti.



Per definire la tabella è stata scelta la frequenza di riferimento di 1000 Hz. Il livello di pressione sonora (in *dB SPL*) che assume ogni *curva isofonica* alla frequenza di 1000 Hz definisce il livello di sensazione sonora in *phon*. Ad esempio, un'onda sonora a 40 *dB*, alla frequenza di 1000 Hz, definisce la sensazione che chiameremo di 40 *phon* (per cui in questo caso il valore in *dB* coincide con quello in *phon*): questa, seguendo la curva di Fletcher e Munson che passa per l'ascissa 1000 Hz e per l'ordinata 40 *dB*, sarà ad esempio la stessa sensazione di intensità che produce all'orecchio dell'ascoltatore una frequenza di 100 Hz emessa con circa 53-54 *dB* di pressione sonora. Analogamente, la sensazione che dà una sorgente di circa 105 *dB* a 20 Hz sarà la stessa di una sorgente di circa 52 *dB* a 3000 Hz, ovvero, una sensazione di 60 *phon*. Questi dati sono sperimentali e possono variare da individuo a individuo. Sono addirittura soggetti a variazioni nel tempo a lungo termine, passando cioè da generazione a generazione (a causa delle abitudini di vita e delle differenze culturali, nonché dei possibili cambiamenti della fisiologia stessa dell'apparato uditivo); sono però molto utili per determinare alcuni fatti di carattere generale. Ad esempio, ed è l'osservazione più importante, si vede bene che bisogna aumentare molto il

“volume” delle frequenze basse perché diano una sensazione sonora simile a quella data dalle frequenze medie, e c’è addirittura un picco di sensibilità nella zona della voce umana, che, come abbiamo già visto, è situata nelle frequenze da circa 2 a circa 5 kHz. Da queste frequenze in poi le curve ricominciano a salire, indicando che la sensibilità alle frequenze alte è minore rispetto a quelle medie. Con l’avanzare dell’età dell’individuo queste curve si modificano con ulteriori perdite di sensibilità verso il basso e verso l’alto (intorno ai 60 anni un uomo non sente al di sopra dei 10 kHz, salvo casi eccezionali). Si noti che la regione al di sopra della prima curva di Fletcher e Munson (quella a 0 phon) e al di sotto dell’ultima (quella a 120 phon) è la cosiddetta area di udibilità dell’orecchio umano (visibile nel grafico del capitolo 1.4).

1.4.2) Dal phon al son

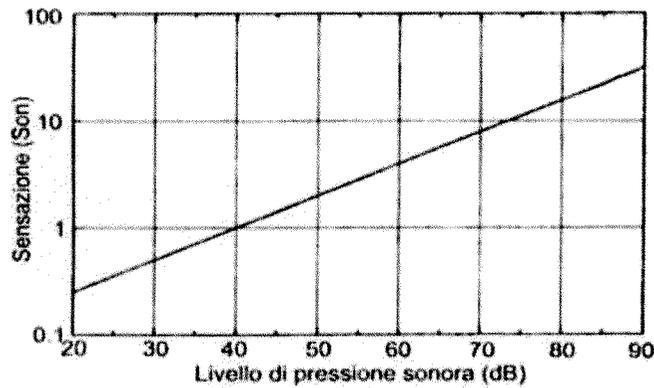
Le curve di isofonia, di cui si è a lungo discusso, descrivono la relazione fra suoni percepiti soggettivamente come ugualmente intensi e il rispettivo valore di intensità oggettiva, ma non consentono di trovare la relazione tra intensità oggettiva ed intensità soggettiva. Non esiste cioè una correlazione diretta fra valori numerici in phon e caratteristiche della percezione auditiva, tranne per il fatto che a valori numerici più elevati corrispondono suoni di livello soggettivo equivalente più intenso. Non possiamo ad esempio dire quante volte è più intenso un suono a 40 phons rispetto ad un suono a 20 phons. Per indicare questa relazione viene quindi utilizzata un’altra unità di misura, il son, che corrisponde alla sensazione sonora indotta da un tono puro a 1 kHz di 40 dB (uguale quindi a 40 phons). La formula per trovare la loudness in son è la seguente:

$$L = 2^{(L_s - 40)/10}$$

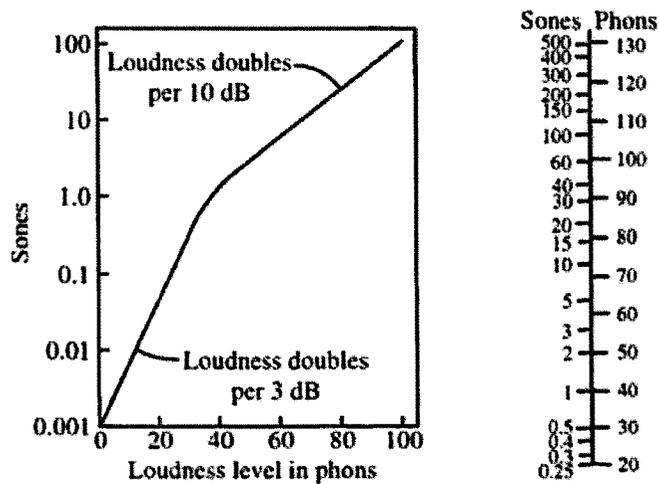
Dove L_s indica il livello di sensazione sonora in phons.

La scala dei son è stata studiata in modo che ad ogni raddoppio di son corrisponda un raddoppio di intensità sonora percepita dal nostro orecchio, che corrisponde anche ad un incremento approssimativo di 10 dB di pressione sonora. La dipendenza della sensazione

dalla pressione è esponenziale, quindi la relazione tra livelli di sensazione sonora e livelli di pressione è lineare.



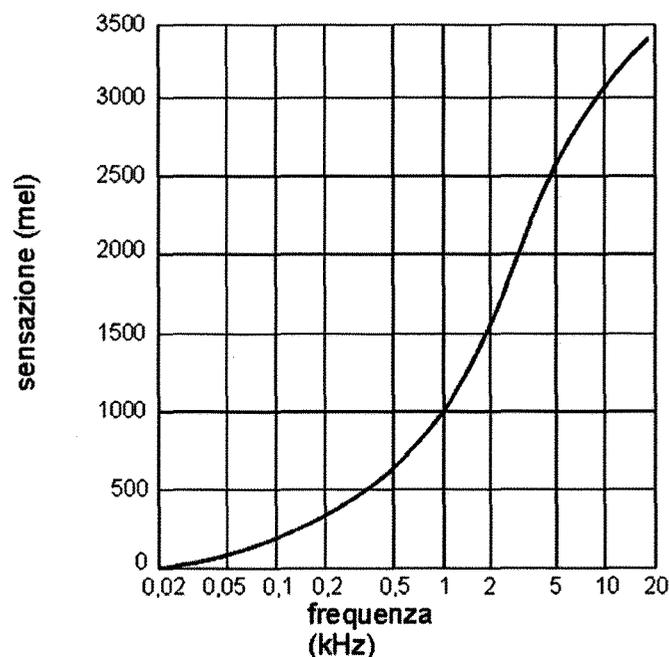
Il seguente grafico mostra la relazione tra la loudness espressa in *phons* e quella espressa in *sones*:



1.4.3) Altezza in funzione della frequenza: il mel

Analogamente a quanto accade per l'intensità, anche l'altezza di un suono non è in relazione strettamente lineare con la frequenza che lo caratterizza. Per questo motivo è stato necessario definire anche per essa un'unità di misura: il *mel*. Per convenzione un tono di 1000 Hz a 40 dB induce una sensazione di altezza pari a 1000 mel. Una sensazione di altezza doppia di quella di riferimento (ottava superiore) corrisponde a 2000 mel; una

sensazione di altezza pari alla metà (ottava inferiore), 500 *mel*. Verificando però i dati ottenuti sperimentalmente si nota che l'effetto di ottava superiore, anziché da 2000 *Hz* è indotto da circa 3100 *Hz* mentre quello di ottava inferiore è indotto da una frequenza di circa 400 *Hz*. Utilizzando sempre un tono di 40 *phon* è possibile dimostrare come al cambiare dell'intensità del tono di stimolazione, mantenuto a frequenza costante, cambia anche la sensazione di altezza, per cui è possibile tracciare sperimentalmente delle *curve di isoaltezza* (identica sensazione di altezza) che dicono come sia necessario variare in più o in meno l'altezza reale dei tono di stimolazione perché l'udito conservi la sensazione di una nota costante variandone l'intensità.



1.4.4) Bande critiche

Sovrapponendo più toni puri, la sensazione soggettiva di intensità del suono risultante è superiore di quella di ciascuno dei toni componenti. Se la frequenza dei due toni puri è molto diversa, allora la sensazione sonora della combinazione dei due toni è data dalla

somma della sensazione di ciascun tono. Se invece la differenza di frequenza dei due toni è ridotta, allora la sensazione sonora indotta sarà inferiore alla somma delle sensazioni sonore dei due toni originali. La *banda critica* è l'intervallo di frequenza nel quale la sensazione sonora indotta dalla somma di due suoni la cui frequenza rientra in questo intervallo è inferiore alla somma delle due sensazioni sonore di ciascun suono. La larghezza delle bande critiche non è costante ma aumenta all'aumentare della frequenza centrale della banda.

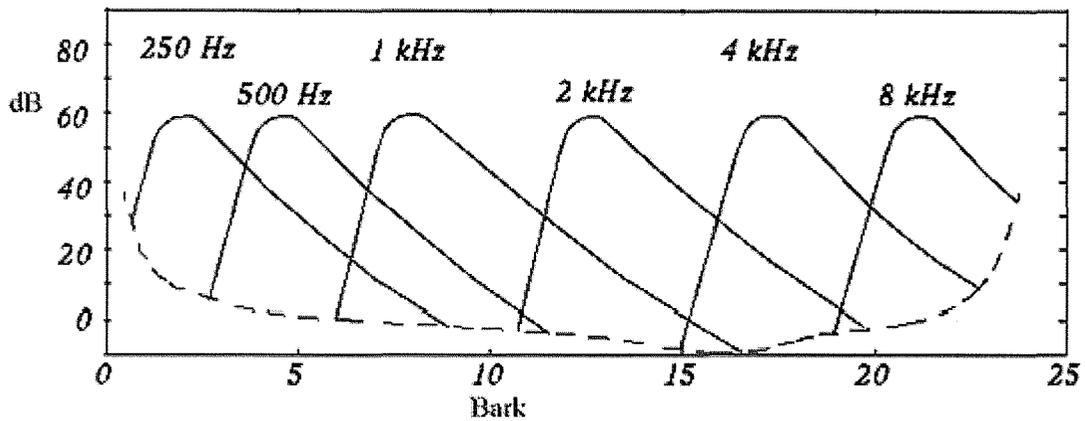
Il concetto di banda critica è importante nella percezione dei *toni complessi*; infatti la sensazione sonora di un tono complesso mantenuto ad un valore costante di intensità, rimane costante fino a quando il suo spettro interessa un intervallo di frequenze inferiori ad una banda critica. Quando le frequenze caratterizzanti lo spettro del tono complesso escono da tale intervallo, la sensazione percepita aumenta con l'aumentare dell'ampiezza dell'intervallo.

Inoltre, se due suoni che facciano parte di una banda critica raggiungono contemporaneamente il sistema uditivo, essi non vengono percepiti come distinti ma come un unico suono di frequenza media rispetto ai 2 originali (Plomp, 1976).

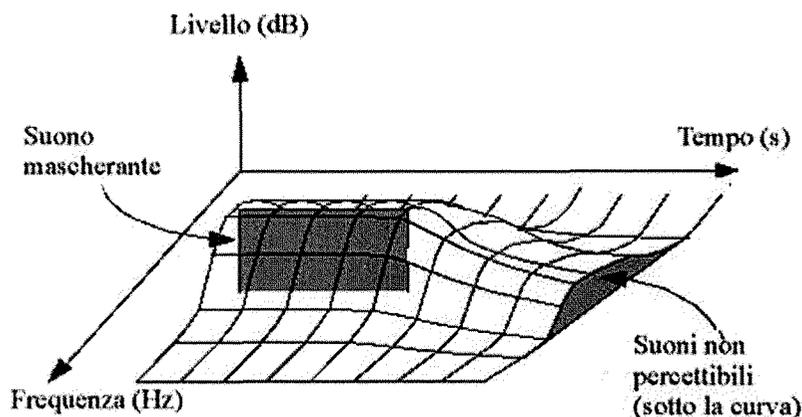
1.4.5) Il mascheramento

Il concetto di banda critica è importante per comprendere effetti acustici come il mascheramento. Il mascheramento è l'innalzamento della soglia di udibilità di un suono meno intenso (*suono mascherato*) ad opera di un suono più intenso (*suono mascherante*). Il livello di mascheramento si misura in *dB*, più precisamente in *dB HL (Hearing Level)*, di aumento della soglia del suono mascherato. L'entità del mascheramento dipende dalla composizione spettrale dei due suoni; esso è maggiore quando i due suoni hanno la stessa frequenza o sono uno multiplo dell'altro. Inoltre, il livello di mascheramento dipende anche dall'altezza del suono mascherante; il mascheramento sarà più intenso nel caso di suoni a bassa frequenza. Questo ragionamento è valido all'interno di una certa gamma di frequenze, ovvero la banda critica di cui si è precedentemente parlato. Il seguente grafico

mostra le aree di mascheramento per suoni a 60 dB alle varie frequenze: l'unità di misura delle ascisse è il *Bark* che corrisponde alla larghezza di una banda critica.



Un altro tipo di mascheramento è il *mascheramento temporale*. Esso è dovuto al fatto che l'udito umano non è in grado di distinguere suoni deboli se raggiungono il nostro orecchio ravvicinati a un tono, simile in frequenza, ma di livello molto alto. In pratica in presenza di un suono forte è necessario che trascorra un dato intervallo di tempo prima di poter distinguere un suono più debole vicino in frequenza. Questo fenomeno è dovuto al fatto che un suono più forte mette in movimento i complicati meccanismi del nostro apparato uditivo per un tempo maggiore rispetto ad un suono debole, è quindi necessario che passi il tempo sufficiente all'orecchio di tornare in posizione di riposo prima di poter sentire il suono seguente senza distorsioni. L'effetto è ben visibile in questo grafico:



Il fenomeno del mascheramento è ben comprensibile in riferimento all'organizzazione tonotopica della coclea. Infatti, come già spiegato nel precedente capitolo, suoni a bassa frequenza eccitano la membrana basilare presso l'elicotrema, mentre suoni ad alta frequenza eccitano la zona vicino alla finestra ovale. In considerazione di quanto detto e del comportamento della coclea, è facile comprendere come il mascheramento sia legato alla parziale sovrapposizione di aree di eccitazione della membrana basilare.

CAPITOLO 2

LOUDNESS: DAI PRIMI STUDI FINO ALLO STANDARD ATTUALE OSI 226

2.1) IL LAVORO DI HARVEY FLETCHER E W. A. MUNSON

Lo studio dell'intensità sonora percepita in un segnale sonoro, cominciò già negli anni '20. Le prime scoperte di carattere tecnologico in ambito di trattamento del suono vennero fatte in quegli anni, questo diede, agli scienziati della Bell Company, i mezzi necessari di intraprendere i loro primi esperimenti in materia di loudness. Ci provarono inizialmente Steinberg e Fletcher nel 1924 e nell'anno seguente: essi cercarono di dimostrare gli effetti di loudness che comportava l'eliminazione di una certa banda di frequenza da un tono complesso e vollero creare una formula per il calcolo dei suoni complessi. La tecnologia era ancora agli albori e tutti i segreti riguardo al funzionamento del nostro apparato uditivo, non erano ancora stati scoperti, per cui i dati ottenuti dai loro esperimenti si rivelarono inconsistenti. Nel decennio seguente, però, diverse nuove scoperte ed un interesse sempre crescente, portarono gli scienziati Harvey Fletcher e W. A. Munson a riaffrontare l'argomento. Nell'Ottobre del 1933 venne pubblicato l'articolo derivante dagli studi svolti dai due scienziati intitolato "Loudness, Its Definition, Measurement and Calculation". Come si evince dal titolo dell'articolo, i due ricercatori hanno voluto dare una definizione del termine "loudness" e dei termini dei vari parametri che rientrano nello studio dell'intensità percepita, hanno esposto le varie tecniche di misurazione ed hanno cercato di creare una formula stabile per il calcolo della loudness.

L'articolo inizia definendo la loudness come un termine psicologico utilizzato per descrivere la grandezza di una sensazione uditiva. I termini utilizzati per indicare un'intensità sonora in musica (come ad esempio "p" per piano o "ff" per molto forte) non

hanno un valore assoluto bensì dipendono dall'esperienza, dalla capacità uditiva e dall'abitudine di chi li usa.

La grandezza della sensazione uditiva dipende dal numero di impulsi nervosi che raggiungono il cervello tramite il canale uditivo. Per questo motivo è evidente che la sensazione uditiva non dipende solamente dall'intensità del segnale che raggiunge il nostro orecchio bensì anche dalla sua composizione fisica (ovvero dalla forma d'onda). L'intensità percepita dipende inoltre dalle condizioni psicologiche e fisiologiche dell'ascoltatore con riferimento a fatica, attenzione, lucidità mentale ecc. Per determinare la loudness prodotta, è quindi necessario definire l'intensità del suono, la sua forma d'onda, il tipo di orecchio ricevente e le condizioni psicofisiche dell'ascoltatore.

2.1.1) Definizioni

Fletcher e Munson, nel loro articolo, definirono alcuni termini, usati fino ad allora spesso erroneamente, necessari per la corretta comprensione dei loro elaborati e con lo scopo di fissare degli standard di linguaggio per futuri studi in questo ambito.

Sound Intensity (intensità sonora): l'intensità sonora è la quantità di energia, veicolata dall'onda sonora, trasmessa in un'unità di tempo attraverso una sezione di area unitaria. Nel caso di un'onda piana o sferica avente un'effettiva pressione sonora P (bars), una velocità di propagazione c (centimetri per secondo) in un mezzo di densità p (grammi per centimetro cubico), l'intensità sonora nella direzione di propagazione è data da:

$$J = P^2 / pc$$

Reference Intensity (intensità di riferimento): come intensità di riferimento per comparare i diversi livelli di intensità sonora è stato scelto il valore di 10^{-16} watts per centimetro quadrato. Questo valore è stato scelto perché rende semplice i lavori di calcolo e perché corrisponde alla soglia di intensità dell'ascoltatore tipo.

Intensity Level (livello di intensità): il livello di intensità di un suono è il numero di *dB* sopra l'intensità di riferimento.

Reference Tone (tono di riferimento): come tono di riferimento è stato scelto un tono puro con una singola frequenza di 1000 cicli al secondo (1 *kHz*). I due ricercatori hanno scelto questo tipo di suono per diversi motivi:

1. è semplice da definire;
2. è spesso usato come standard di riferimento per il *pitch* (altezza);
3. il suo utilizzo rende la formula matematica più semplice;
4. il suo range di intensità uditive è più ampio di qualsiasi altro tipo di suono;
5. la sua frequenza è circa a metà del range delle frequenze udibili.

Loudness Level (livello di intensità percepita): si ottiene quindi regolando il livello di intensità del tono di riferimento fino al punto in cui risulterà percepito di eguale intensità, dal tipico ascoltatore, rispetto al tono sconosciuto. Il solo modo per determinare l'*ascoltatore tipo* è quello di sottoporre a test acustici un numero di persone che hanno un udito normale; l'*ascoltatore tipo* sarà colui che da come risultato la media esatta dei risultati degli altri ascoltatori esaminati.

Nel determinare il livello di loudness con paragone al tono di riferimento, ci sono due classi generali di suoni per le quali sono necessarie misurazioni: quelli stazionari (come ad esempio il ronzio dei macchinari) e quelli la cui intensità percepita varia (come il rumore stradale, la musica, una conversazione, ecc.). Un suono stazionario può essere rappresentato da un numero finito di toni puri chiamati *componenti*; il valore dei componenti, nella posizione di ascolto dove si vuole ottenere il livello di loudness, è dato dai valori di intensità di ogni singolo componente in quella posizione.

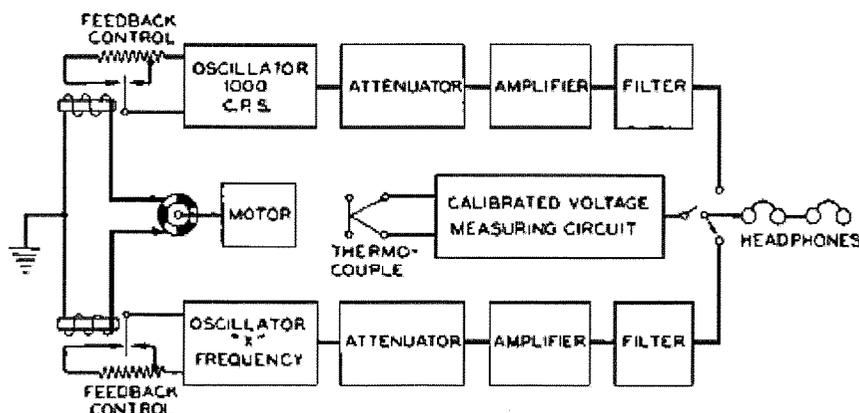
In aggiunta alla frequenza ed alla grandezza dei componenti di un suono è necessario conoscere la posizione e l'orientazione della testa dell'*ascoltatore* rispetto alla posizione della sorgente sonora ed anche se per l'ascolto è stato usato un orecchio solo oppure entrambe.

2.1.3) Metodo sperimentale per la misurazione del livello di loudness di un suono statico

Il metodo utilizzato da Fletcher e Munson, nei loro esperimenti, per ottenere una misura del livello di loudness di un suono, consisteva nell'ascolto alternato di questo suono con un tono di riferimento a 1000 Hz e nell'aggiustare il livello di intensità di quest'ultimo fino a che il volume percepito non risultasse uguale al suono da analizzare. Quando si ottiene questa condizione il tono di riferimento sarà di L dB e quindi si dirà che il tono in esame avrà un livello di loudness di L dB. Quando le caratteristiche del tono da misurare risultano simili a quelle del tono di riferimento, il paragone è semplice e la misurazione facile da effettuare. Per altri suoni invece, i diversi fattori che entrano in gioco durante il processo di ascolto diventano importanti ed è quindi necessario ricorrere ad un metodo sperimentale che darà risultati tipici di un orecchio medio e di un soggetto in normali condizioni fisiologiche e psicologiche.

Per determinare se due suoni hanno lo stesso volume, è necessario basarsi sui giudizi di un osservatore che vengono influenzati, oltre che dai meccanismi dell'orecchio, anche dai processi mentali associati all'ascolto. Questo porta ad una notevole varietà nei risultati degli ascoltatori che costringono il ricercatore a lavorare su basi statistiche. Il metodo dello *stimolo costante*, nel quale l'ascoltatore deve scegliere tra due suoni a intensità fissa quello che percepisce in maniera maggiore, sembra il più adatto al controllo dei vari fattori che interferiscono coi giudizi quando si utilizzano più ascoltatori contemporaneamente. Con questa tecnica l'ascoltatore è limitato esclusivamente a dare un parere di loudness; questo è molto importante perché è stato constatato che l'utilizzo delle apparecchiature o la partecipazione a qualsiasi altra attività che non sia quella puramente di ascolto, influenzano negativamente i risultati dei test, che sono ulteriormente aggravati dall'utilizzo continuo degli stessi ascoltatori per un lungo periodo di tempo. Il controllo della fatica, gli effetti di memoria e l'associazione del giudizio di un osservatore con i risultati del test o con il giudizio di un altro ascoltatore possono essere rigidamente mantenuti con questo metodo.

Di seguito è illustrato il circuito utilizzato dai due ricercatori (Bell Company, 1933) per la generazione ed il controllo del tono di riferimento e del tono da analizzare:



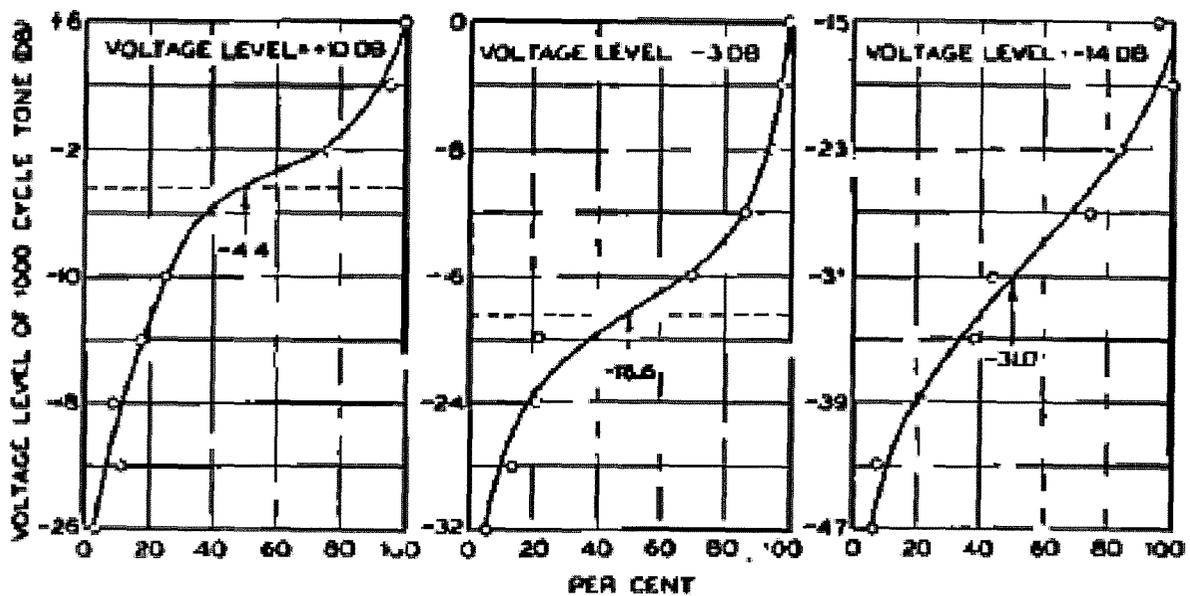
Due oscillatori sono stati utilizzati per generare toni puri. Con l'utilizzo del circuito di misura del voltaggio e con l'attenuatore, il livello di voltaggio mandato in cuffia poteva essere determinato. L'attenuatore, che era calibrato in *dB*, è stato fissato in modo che la tensione di misurazione di 1 *volt* indicata venisse impressa sul ricevitore. La differenza, poi, tra questa impostazione e le altre impostazioni è il livello di voltaggio. Per ottenere il livello di intensità di un suono dobbiamo conoscere la taratura dell'apparecchio ricevente.

Gli ascoltatori vennero posizionati in una cabina insonorizzata. Venne chiesto loro solamente di ascoltare e di operare una scelta, operazione che non può essere vista dagli altri ascoltatori per evitare che il giudizio di uno influenzi anche il giudizio di qualche altro. Agli ascoltatori venne proposto un suono ed immediatamente successivamente il tono di riferimento, ognuno dei quali per la durata di un secondo. Dopo una pausa di un secondo la sequenza venne ripetuta e venne chiesto loro di indicare quale dei due suoni era percepito maggiormente, se quello da testare oppure quello di riferimento, e di operare una scelta con un semplice *switch*. I livelli vennero poi cambiati ed il test ripetuto.

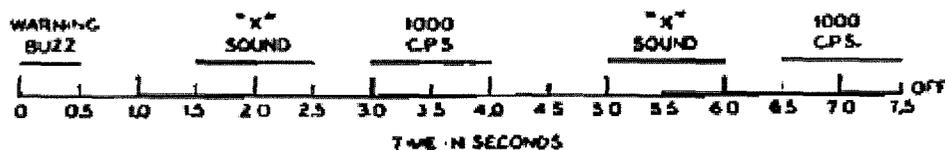
Nella tabella seguente si possono notare i risultati di un test fatto a tre ascoltatori per un tono di 125 *Hz* a tre differenti livelli:

Obs.		+6	+2	-2	-6	-10	-14	-18	-22	-26
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	0	0	0	0	0	0
level=	CK	+	+	+	+	+	0	0	0	0
+9.8db	AS	+	+	+	+	0	0	0	0	0
	DH	+	+	0	0	+	0	0	0	0
	CK	+	+	+	+	0	0	0	0	0
	AS	+	+	+	0	0	0	0	0	0
	DH	+	+	0	0	0	0	0	0	0
Obs.		0	-4	-8	-12	-16	-20	-24	-28	-32
125	CK	+	+	+	+	0	+	+	0	0
c.p.s.	AS	+	+	+	+	+	0	0	0	0
Volt.	DH	+	+	+	+	0	0	0	0	0
level=	CK	+	+	+	+	+	+	+	0	0
-3.2db	AS	+	+	+	+	+	+	+	0	0
	DH	+	+	+	0	+	0	+	0	0
	CK	+	+	+	+	+	+	0	0	0
	AS	+	+	+	+	+	0	0	0	0
	DH	+	+	+	0	+	0	0	0	0
Obs.		-15	-19	-23	-27	-31	-35	-39	-43	-47
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	+	+	+	0	0	0
level=	CK	0	+	+	+	+	+	0	0	0
-14.2	AS	+	+	+	+	0	+	0	0	0
db	DH	+	+	0	+	0	0	+	0	0
	CK	+	+	0	+	+	+	0	0	0
	AS	+	+	0	0	+	+	0	0	0
	DH	+	+	0	0	0	0	+	0	0

Nel grafico, lo 0 indica che l'ascoltatore ha percepito di maggiore intensità il tono a 125 Hz mentre il segno + indica che il tono di riferimento è stato percepito maggiormente. Pareri di uguaglianza tra i due toni non erano permessi. Il valore all'inizio di ogni colonna indica il livello di voltaggio del tono di riferimento mandato in cuffia, che è il numero di dB di differenza da 1 volt (+ se è sopra e - se è sotto) mentre quelli a lato sono i valori del tono da testare. Successivi test sono stati scelti a caso tra le ventisette combinazioni possibili dei livelli mostrati, questo per ridurre la possibilità di effetti di memoria. I livelli erano stati selezionati così che il tono di riferimento suonasse sia di maggiore che minore intensità rispetto al tono in esame e la media dei loro giudizi venne presa come punto di "equal loudness". L'esperimento venne eseguito su un totale di undici ascoltatori. La freccia all'interno del seguente grafico indica il livello medio al quale il tono di riferimento a 1000 Hz è stato percepito di eguale intensità, dal gruppo in esame, rispetto al tono di 125 cicli/sec.

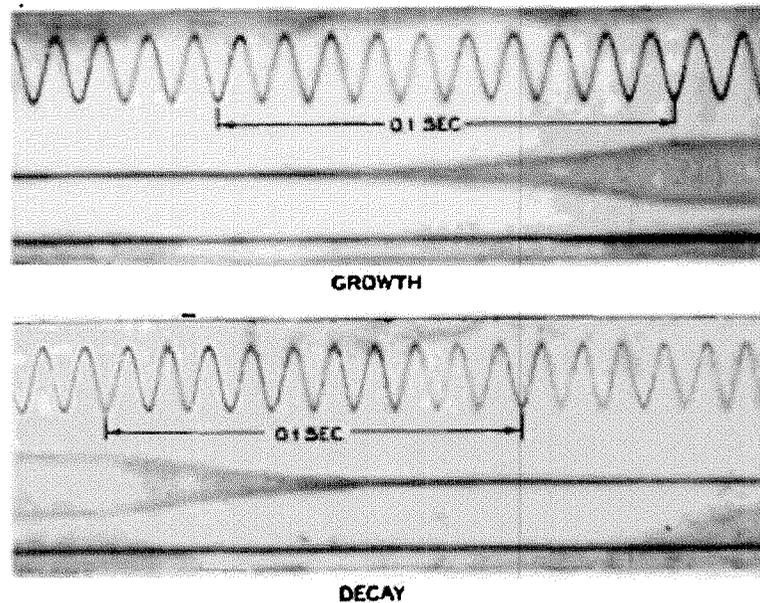


Alcuni metodi per minimizzare la fatica fisica e mentale dell'ascoltatore sono stati adottati durante i test. L'esperimento è stato strutturato su sessioni di 15 minuti l'una, ognuna delle quali dava la possibilità di effettuare circa 15 test. Il massimo numero di ascolti permessi in un giorno era 150. Per evitare l'affaticamento dell'orecchio, i suoi proposti agli ascoltatori erano di breve durata e proposti nel seguente ordine:



Il suono presentato doveva essere sufficientemente corto da non affaticare l'orecchio dell'ascoltatore e nello stesso tempo sufficientemente lungo a raggiungere la sua piena capacità di loudness. I due toni inoltre sono stati presentati ad intervalli di tempo studiati per garantire un corretto paragone e lasciare un sufficiente periodo di riposo all'orecchio. E' stato rilevato che ad alti livelli occorrono 0.3 secondi prima che il tono raggiunga la sua piena loudness e che, se presentato per un periodo di esposizione maggiore di un secondo, può affaticare l'orecchio dell'ascoltatore. Per evitare poi gli effetti di distorsione che si creano nell'orecchio quando un suono ad alto livello viene interrotto improvvisamente, il

circuito di controllo fermava ed avviava i suoni in maniera graduale. Il periodo di crescita e di decadimento è stato settato a 0.1 secondi:



Rilevamenti della soglie vennero fatti prima e dopo i test. Vennero eseguiti con gli stessi circuiti utilizzati per l'esperimento di loudness, disattivando l'oscillatore a 1000 Hz ed attenuando lentamente il secondo tono fino a quando non risultasse più udibile, per poi rialzarlo sopra alla soglia di udibilità. Gli ascoltatori dovevano segnalare quando non gli era più possibile udire il suono e quando invece il suono tornava udibile. La media di queste due condizioni venne presa come soglia di udibilità.

Un'ulteriore analisi delle armoniche generate dalle cuffie e dagli altri apparati venne fatta per essere sicuri della purezza del suono che doveva raggiungere l'orecchio. Le cuffie erano di tipo elettrodinamico e venne scoperto che producevano armoniche dell'ordine dei 50 dB sotto la fondamentale. A livelli molto alti, le distorsioni causate dai filtri erano maggiori di quelle generate dalle cuffie; in ogni caso il livello di loudness di ogni armonica era di 20 dB o più sotto il livello della fondamentale. Esperienze con toni complessi hanno dimostrato che in condizione come queste, il contributo delle armoniche sulla loudness totale è insignificante.

Il metodo di misurazione della loudness che è stato descritto venne utilizzato per una larga varietà di suoni ed ha spesso dato risultati soddisfacenti.

2.1.4) Confronto con il lavoro di B. A. Kingsbury

Nel 1926 B. A. Kingsbury, anch'egli ricercatore della Bell Company, fece un esperimento analogo a quello svolto da Fletcher e Munson alcuni anni dopo. La loudness di undici toni puri venne studiata regolando il voltaggio applicato ad un ricevitore telefonico fino a che questi toni non risultassero di uguale intensità rispetto ad alcuni livelli fissati di un tono a 700 Hz. La media dei risultati di 22 ascoltatori (11 maschi ed 11 femmine) venne impostata come curve di isofonia, attraverso l'area di normale sensazione uditiva in termini di pressione RMS all'interno del canale uditivo, in funzione della frequenza. Venne scoperto che se le ampiezze dei toni puri vengono aumentate in proporzioni uguali, la loudness dei toni a bassa frequenza cresce più rapidamente di quelle dei toni ad alta frequenza.

Come unità di loudness venne utilizzato il minimo incremento di loudness percepibile di un tono a 1000 Hz. Come valore assoluto esso varia da livello a livello ma tende a rimanere costante nella gamma ordinaria di loudness. Questa unità tiene conto del carattere soggettivo della loudness.

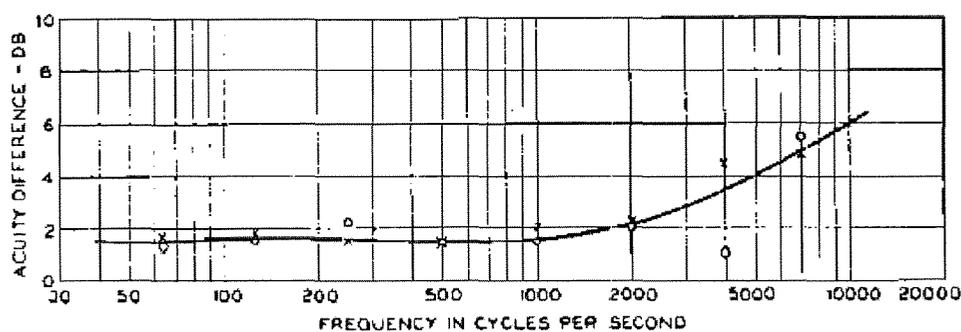
La variabilità dei dati dai quali venne calcolata la media venne separata in un fattore che esprimesse la dissomiglianza tra le orecchie ed un altro che esprimesse gli errori dei giudizi degli ascoltatori. Non vi è stato un livello al quale le variazioni sono state minime. Si è dimostrato poi che le differenze tra le orecchie causarono maggiori variazioni rispetto agli errori commessi dagli ascoltatori e che il sesso dell'ascoltatori non ha un valore rilevante.

Fletcher e Munson fecero un paragone tra i risultati ottenuti dai loro esperimenti e quelli ottenuti da Kingsbury qualche anno prima. Da sottolineare è il fatto che gli ascoltatori negli esperimenti di Fletcher e Munson, ascoltavano i toni con entrambe le orecchie, mentre nel caso di Kingsbury, l'ascolto era monoaurale. Importante è anche il fatto che i livelli dei toni utilizzati negli esperimenti, erano espressi come numero di dB sopra la

soglia media ottenuta con un singolo ricevitore. Per entrambe i motivi, un paragone diretto tra i due lavori non può essere fatto. Tuttavia, nel corso del lavoro di Fletcher e Munson, due serie di esperimenti vennero fatti e questo diede la possibilità di convertire i dati ottenuti da Kingbury per poter paragonare i due lavori. E' stato inizialmente scoperto che se un ascoltatore viene sottoposto all'ascolto del tono di riferimento e del tono di differente frequenza, con entrambe le orecchie e ritiene che i due suoni abbiano la stessa intensità, darà lo stesso parere anche ascoltando il suono con un singolo ricevitore (il voltaggio in cuffia rimane invariato). Quando vengono ascoltati i due suoni con un singolo ricevitore, la loudness di entrambe risulterà dimezzata; di conseguenza se i due toni avevano la stessa intensità sonora percepita con l'ascolto binaurale, essi avranno la stessa loudness anche con l'ascolto monoaurale.

La seconda serie di dati riguarda la differenza di soglia tra l'ascolto monoaurale e binaurale. E' noto che per un individuo, le due orecchie hanno differenti acutezze auditive. Di conseguenza, quando si ascolta con entrambe le orecchie, la soglia è determinata dall'orecchio con la maggior acutezza (l'orecchio migliore).

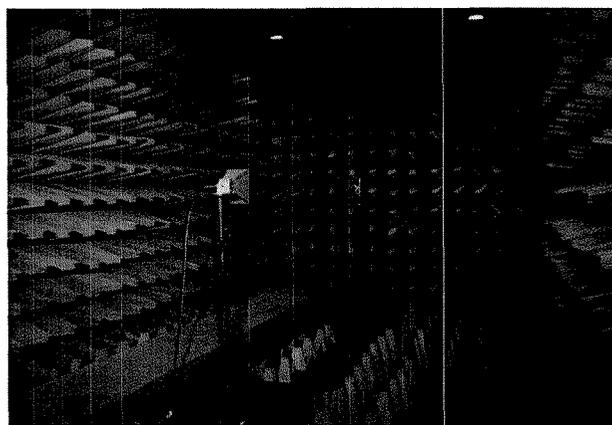
La seguente curva mostra la differenza nel livello di soglia tra la media del miglior orecchio di un ascoltatore e la media di tutte le orecchie:



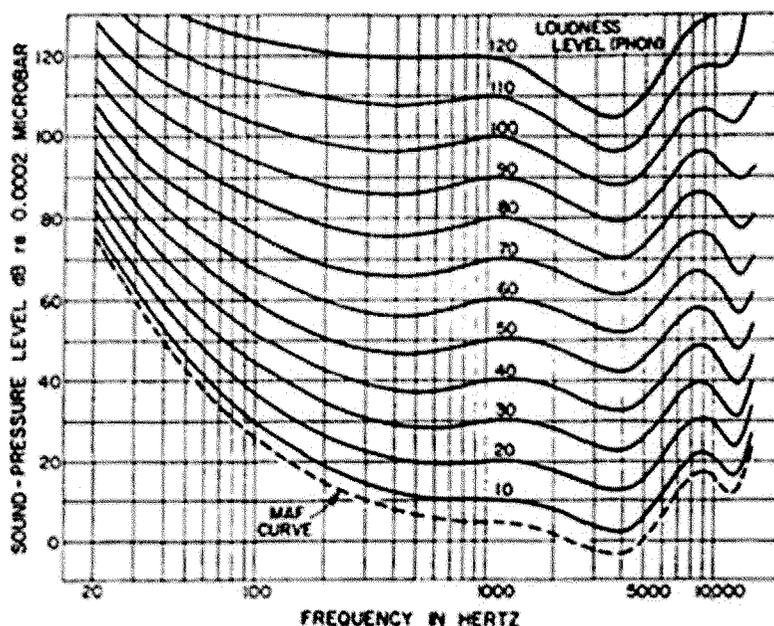
Se la differenza di acutezza quando si ascolta con un orecchio rispetto a quando si ascolta con entrambe è determinata solo dall'orecchio migliore, allora la precedente curva indica questa differenza. Tuttavia, alcuni test sperimentali eseguiti dai due ricercatori per testare l'acutezza del miglior orecchio rispetto a quella di entrambe, hanno mostrato che nel secondo caso il risultato è leggermente migliore. Nonostante ciò, per le frequenze più alte,

2.2) GLI STUDI DI D. W. ROBINSON E R. S. DADSON

In seguito ai lavori di ricerca e sperimentazione svolti dai due ricercatori della Bell Company, altri esperti del settore si cimentarono nello studio dell'intensità sonora percepita. I primi che ci provarono furono Churcher e King nel 1937 ed i risultati mostrarono differenze apprezzabili rispetto a quelli ottenuti dagli sperimentatori originali (Fletcher e Munson). Tuttavia, fu notevolmente più rilevante il lavoro svolto da D. W. Robinson e R. S. Dadson che nel 1956 scoprirono le curve isofoniche che, come vedremo meglio in seguito, vennero adottate come standard dalla normativa ISO 226. Essi, a differenza di Fletcher e Munson che usarono suoni riprodotti in cuffia per svolgere i test, utilizzarono suoni puri riprodotti direttamente da un altoparlante in *campo libero*. Con questo termine si intende che il suono generato da una sorgente si propaga in un mezzo illimitato, privo di discontinuità od ostacoli. Condizioni di questo tipo sono ovviamente solo un'idealizzazione delle reali condizioni in cui un suono si propaga. Nel caso però in cui ci trovassimo in uno spazio all'aperto, con condizioni atmosferiche stabili e omogenee, e non ci siano ostacoli in una zona sufficientemente ampia attorno alla sorgente, possiamo approssimare la situazione con quella di campo libero. In alternativa, è possibile ricreare le condizioni di campo libero a livello di laboratorio operando nella cosiddetta *camera anecoica*, in cui tutte le superfici assorbono completamente le onde sonore che le investono e l'unico suono esistente all'interno è quello prodotto dalla sorgente, per assenza di suono riflesso dalle superfici.



Rispetto all'audiogramma normale di Fletcher e Munson le differenze si hanno soprattutto nei campi di frequenza prossimi a 4 kHz e 8 kHz, questo dovuto al differente metodo utilizzato negli esperimenti per l'ascolto dei toni (in cuffia per Fletcher e Munson e con un altoparlante per Robinson e Dadson). In entrambe le famiglie di curve si nota comunque una zona di scarsa sensibilità sonora in corrispondenza delle basse frequenze, mentre intorno ai 3-4 kHz si ha una zona di massima sensibilità. Ciò significa che a basse frequenze la sensazione sonora aumenta rapidamente all'aumentare del livello di pressione sonora; ad alti valori di frequenza le isofoniche sono quasi parallele e la sensazione aumenta meno rapidamente, il che è un indicatore di non linearità. Una novità introdotta dai due scienziati è la curva isofonica a 4,2 phon, indicata con MAF (Minimum Audible Field) che rappresenta, per ogni frequenza, il minimo livello di pressione sonora per cui un suono è udibile, e corrisponde alla soglia uditiva. La norma ISO 226, come vedremo tra poco, è stata revisionata negli ultimi anni, sulla base di studi condotti negli anni ottanta e novanta in numerosi laboratori. Nel seguente grafico sono tracciate le curve isofoniche di Robinson e Dadson:



2.3) LO STANDARD ISO 226

Le curve di uguale intensità derivate utilizzando cuffie auricolari sono valide solo nel caso speciale di uno stimolo laterale; tuttavia questo non è il modo in cui normalmente l'uomo percepisce i suoni. Nella vita reale i rumori ci investono come onde piane se provengono da una sorgente sufficientemente distante. Nel caso in cui questi si trovi direttamente davanti all'ascoltatore entrambe le orecchie ricevono la stessa intensità.

Nel caso in cui i suoni siano emessi a frequenze inferiori a 1 *kHz* il suono che entra nel canale auricolare è parzialmente ridotto da un effetto di mascheramento della testa, che dipende, in buona parte, dalla riflessione indotta dal padiglione auricolare. Suoni che non provengano da una sorgente centrata sono soggetti maggiormente a questo effetto combinato di mascheramento e riflessione in un orecchio piuttosto che nell'altro. L'insieme dei due effetti è quantificato da insiemi di curve in uno spazio tridimensionale note con il nome di funzioni di trasferimento legate alla testa (HRTF).

Gli standard ISO sono basati su uno stimolo frontale e centrale che è considerato preferibile nel caso in cui si vogliano tracciare i diagrammi di uguale intensità sonora.

Robinson e Dadson avevano utilizzato, nel loro esperimento, degli altoparlanti e, per lungo tempo, le differenze con le curve originali (di Fletcher e Munson) furono spiegate con il diverso apparato di produzione dei suoni. Tuttavia l'ISO, nella sua relazione di accompagnamento allo standard, dichiara di aver utilizzato cuffie adattate per compensare la differenza, senza tuttavia riferire i dettagli di come si sia ottenuta questa compensazione.

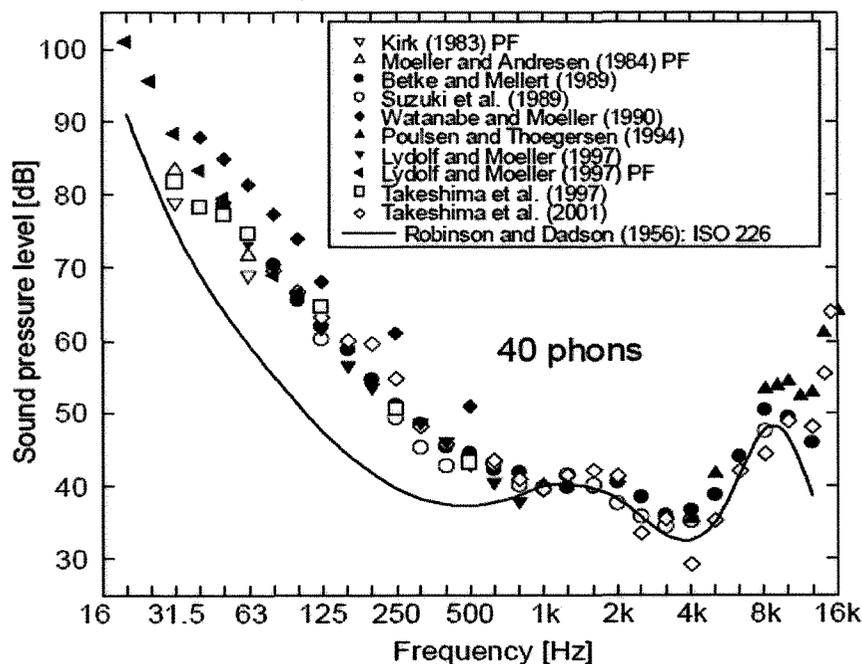
Come già accennato, le curve isofoniche adottate per molti anni dallo standard ISO, furono quelle nate dagli studi svolti in Inghilterra negli anni '50 da Robinson e Dadson. Alcune ricerche pubblicate nella metà degli anni '80 hanno però dato risultati significativamente diversi rispetto a quelli ottenuti da Robinson e Dadson. A causa delle discrepanze fra i risultati più vecchi e quelli più recenti l'International Organization for Standardization (ISO) ha recentemente rivisto le curve definite nello standard ISO 226. Il processo di

revisione fece seguito ad alcune raccomandazioni pubblicate dall'istituto di comunicazione elettrica dell' Università di Tohoku (Giappone).

Al termine del lavoro furono prodotte nuove curve combinando i risultati di diversi esperimenti condotti in Giappone (che fornì il maggior contributo con circa il 40% dei dati), Germania, Danimarca, Regno Unito, e Stati Uniti d'America e fu quindi definito il nuovo standard ISO 226:2003. La relazione al nuovo standard è dedicata anche a discutere le differenze sorprendenti con la versione precedente delle curve e viene notato, fra l'altro, che i dati attuali mostrano migliore accordo con le curve originali di Fletcher e Munson piuttosto che con quelle prodotte nel 1956 da Robinson e Dadson. Queste ultime sembrano differire fino a un massimo di 10 - 15 dB da quelle standardizzate, specialmente nella regione di bassa frequenza

Di rilevanza fondamentale per la revisione dello standard fu lo studio svolto dall'istituto di comunicazione elettrica della Università di Tohoku, sotto la direzione di Yôiti Suzuki.

Dal seguente grafico si può notare come gli studi compiuti negli anni '80 e '90 da diversi ricercatori, abbiano dato risultati significativamente diversi, soprattutto alle basse frequenze, rispetto ai lavori di Robinson e Dadson.



Le differenze visibili nel grafico sfiorano i 15 dB; una tale differenza non ha solo importanza teorica bensì ha anche implicazioni pratiche. Ad esempio, l'attuale *curva di ponderazione A*, utilizzata nei fonometri, è basata sulla curva isofonica a 40 *phons* di Fletcher e Munson e non sul modello di Robinson e Dadson.

2.3.1) Revisione dello standard: le nuove curve di Yôiti Suzuki

Nella seguente tabella sono indicati, in ordine cronologico, tutti gli studi svolti sulle curve isofoniche dal 1927 fino al 2001.

Table I: Chorography of studies on equal-loudness contours

Year	Researchers	Listening condition	Number of subjects(age)	Method	Ref. tone freq. (level)	Test tone freq.
1927	Kingsbury[4]	Earphone	22 (unspecified)	MA	700 (fix)	Hz 60-4000 Hz
1933	Fletcher-Munson[3]	Earphone with correction	11 FF (unspecified)	CS	1 kHz (variable)	62-16000 Hz
1937	Churcher-King[5]	FF	10 (unspecified)	CS	1 kHz (fix)	54-9000Hz
1955	Zwicker-Feldtkeller[6]	Earphone with equalizer	8 FF (unspecified)	Modified Békésy	1 kHz (fix)	50-16000Hz
1956	Robinson-Dadson[1]	FF	90(16-63)/30 (ave. 30)	CS	1 kHz (variable)	25-15000Hz
1972	Whittle <i>et al.</i> [7]	PF	20 (ave. 20)	CS	higher freq (fix)	3.15-50 Hz
1983	Kirk[8]	PF	14 (18-25)	RMLSP	63 Hz (fix)	2-63 Hz
1984	Møller-Andresen[9]	PF	20 (18-25)	RMLSP	63 Hz (fix)	2-63 Hz
1989	Betke-Mellert[10]	FF	13-49 (17-25)	CS	1 kHz (fix)	50-12500Hz
1989	Suzuki <i>et al.</i> [11]	FF	9-32 (19-25)	CS	1 kHz (fix)	31.5-16000Hz
1990	Fastl <i>et al.</i> [12]	FF	12 (21-25)	CS	1 kHz (fix)	100-1000Hz
1990	Watanabe-Møller[13]	FF	10-12 (18-30)	Bracketing	1 kHz (fix)	25-1000Hz
1994	Müller-Fichtl[14]	Open headphones	8(21-25)	CP	-----	62.5-10000Hz
1994	Poulsen-Thøgersen[15]	FF	29 (18-25)	Bracketing	1 kHz (fix)	1000-16000Hz
1997	Lydolf-Møller[16]	FF	27 (19-25)	RMLSP	1 kHz (fix)	50-1000Hz
		PF	27 (19-25)	RMLSP	100 (fix)	Hz 20-100 Hz
1997	Takehima <i>et al.</i> [17]	FF	9-30 (19-25)	CS	1 kHz (fix)	31.5-12500Hz
1999	Bellmann <i>et al.</i> [18]	FF	12 (unspecified)	Adaptive 1up-1down	1 kHz (fix)	100-1000Hz
		PF	12 (unspecified)	Adaptive 1up-1down	100 (fix)	Hz 16-160 Hz
2001	Takehima <i>et al.</i> [19]	FF	7-32 (18-25)	RMLSP	1 kHz (fix)	50-16000Hz

Lo scopo di Suzuki, ricercatore dell' Università di Tohoku, e del suo team è quello di creare nuove e più affidabili curve isofoniche da inserire nel nuovo standard ISO 226. Per fare ciò, lui ed il suo team analizzarono tutti gli esperimenti fatti fino ad allora, prendendo in considerazione le condizioni di ascolto, il numero e l'età degli ascoltatori, i valori del tono di riferimento e quelli del tono di test. Dei 19 esperimenti elencati in tabella, il team di ricerca dell'università di Tohoku decise di scartarne 3 dalla raccolta di dati: 1. i lavori di Kingbury non vennero presi in considerazione poiché vennero svolti con ascolto monoaurale tramite ricevitore telefonico, metodo non adatto alla stesura delle nuove curve; 2. vennero esclusi anche gli esperimenti fatti da Whittle in quanto fece misurazione di loudness solamente a 3.15, 6.3, 12.5 e 25 Hz e con toni di riferimento settati a 6.3, 12.5, 25 e 50 Hz. Nessuna misurazione venne fatta con un tono di riferimento a 1 kHz, a causa di ciò i livelli ottenuti da questi studi non possono essere espressi direttamente in *phon*; 3. infine, vennero scartati i lavori di Müller e Fichtl, i quali usarono una tecnica che non diede rilevanza agli effetti di contesto nello svolgimento dell'esperimento, come la spaziatura e la frequenza di presentazione dello stimolo.

Quattro studi, quelli di Fletcher and Munson, Churcher and King, Zwicker and Feldtkeller e Robinson and Dadson, offrono un set completo di curve isofoniche mentre i rimanenti lavori mostrati in tabella riportarono solo misurazioni dei livelli di uguale intensità percepita. Queste quattro serie di curve sono state prese in considerazione come curve isofoniche *classiche*, mentre i dati ottenuti dai lavori a partire dal 1983 vennero considerati dati sperimentali *recenti*. Nonostante qualche differenza tra i dati *recenti*, è chiaro che questi dati seguono un trend comune. Facendo un paragone, nessun set di curve *classiche* coincide accettabilmente con i dati *recenti* sull'intero range di frequenze; per questo motivo venne deciso di strutturare le nuove curve isofoniche in base ai dati *recenti*.

E' opportuno disegnare la soglia di udibilità come un limite inferiore di udibilità in un disegno di curve isofoniche, e sarà utile per stimare le nuove curve descritte in seguito. Nella maggior parte delle ricerche sulle curve isofoniche, le soglie di udibilità venivano misurate allo stesso tempo in cui veniva eseguito l'esperimento. La soglia di udibilità, generalmente, coincide tra gli studi *recenti* e si adatta bene alla curva di soglia di Robinson e Dadson, per questo motivo il team di ricerche scelse di utilizzare proprio questa curva.

Inoltre, soglie di udibilità di toni puri in campo aperto, vennero misurate con esperimenti compiuti da altri ricercatori non presenti in tabella (Teranishi, Brinkmann, Vorländer e Poulsen e Han). Anche i dati di questi studi, assieme a quelli presenti in tabella dal 1983 in poi, vennero usati per calcolare le nuove curve isofoniche.

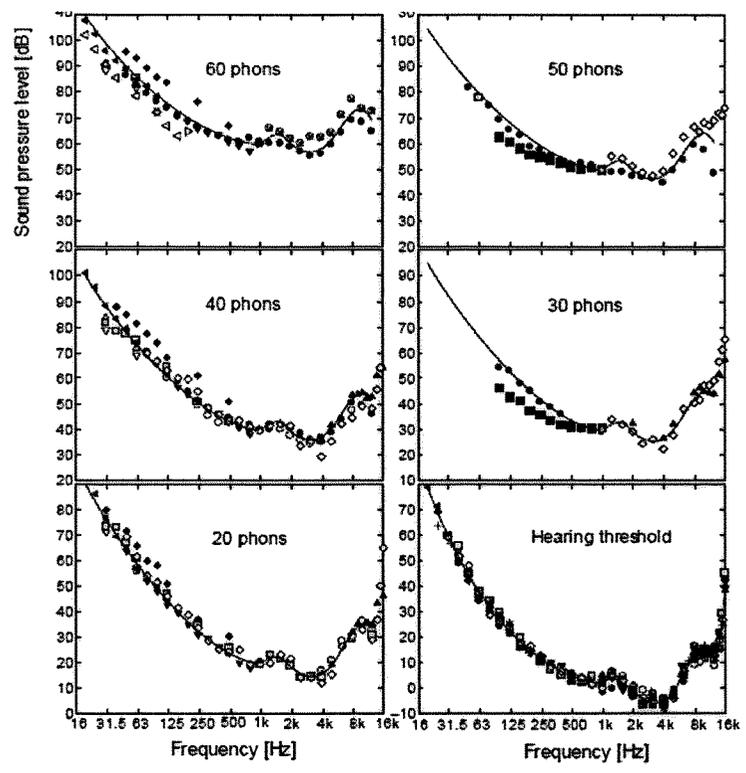
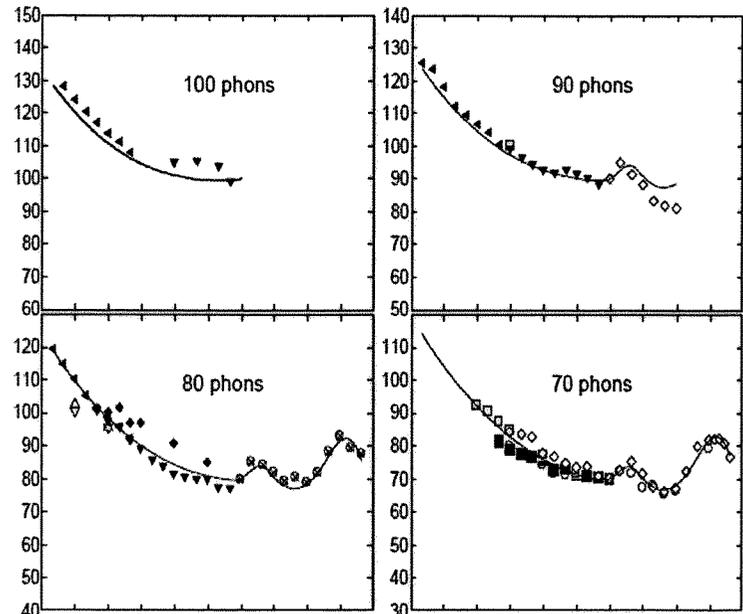
Per disegnare le curve isofoniche dai dati sperimentali, un'interpolazione lungo l'asse delle frequenze è necessaria, perché tutti i dati sperimentali sono proposti distintamente per singole frequenze e livelli di loudness. Inoltre, siccome questi dati mostrano differenze tra i diversi soggetti e studi, è richiesta una "levigatura" appropriata delle curve. Utile per la stesura delle curve è anche una funzione di loudness, ossia una funzione che rappresenti la crescita di loudness in funzione del livello di intensità sonora. Ad oggi, diversi modelli sono stati proposti per descrivere la crescita di loudness in funzione della pressione acustica. Tutte queste funzioni sono essenzialmente basate sulla legge di potenza di Stevens. Questa legge afferma che l'intensità soggettiva (cioè la sensazione) aumenta proporzionalmente all'intensità effettiva dello stimolo elevata ad una potenza. Questa potenza varia a seconda dei sensi, nel caso dell'udito è 0,3. Infatti:

$$L = K I^{0,3}$$

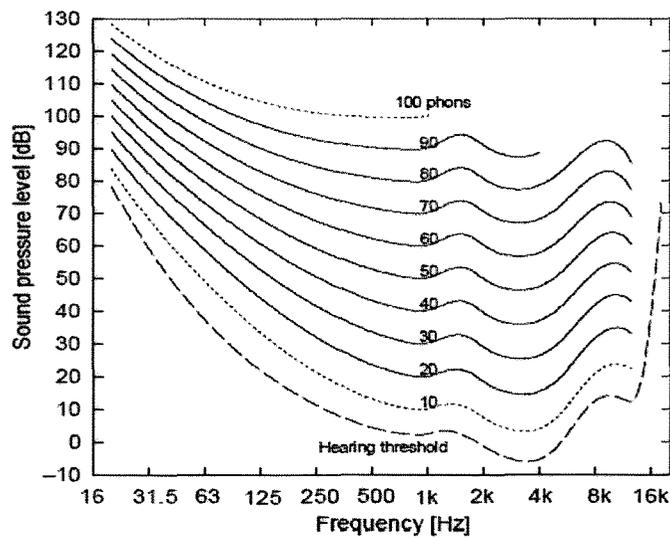
dove L è la sensazione sonora ed I è l'intensità sonora dello stimolo.

Quale tra queste funzioni meglio descriva la crescita di loudness a bassi livelli è ancora un'importante questione non risolta, perché dati recenti mostrano che il livello di differenza tra la soglia dell'udibilità ed i 20 *phons* è generalmente più largo di quello tra i 20 ed i 40 *phons* in regioni di bassa frequenza.

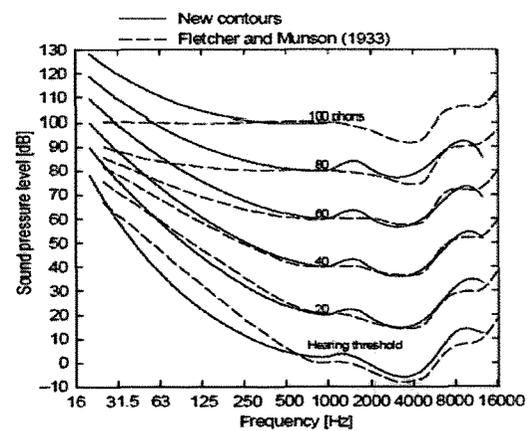
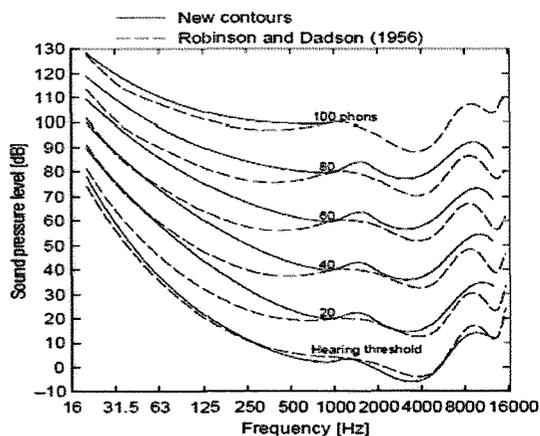
- ▽ Kirk (1983) PF
- △ Moeller and Andresen (1984) PF
- Betke and Mellert (1989)
- Suzuki et al. (1989)
- Fastl et al. (1989)
- ◆ Watanabe and Moeller (1990)
- ▲ Poulsen and Thøgersen (1994)
- ▼ Lydolf and Moeller (1997)
- ◄ Lydolf and Moeller (1997) PF
- Takeshima et al. (1997)
- ▷ Bellmann et al. (1999)
- ◁ Bellmann et al. (1999) PF
- ◇ Takeshima et al. (2001)
- ⊗ Takeshima et al. (2002)
- + Robinson and Dadson (1956) HT
- ▲ Teranishi (1965) HT
- × Brinkmann (1973) HT
- * Vorlander (1991) HT
- * Poulsen and Han (2000)
- Estimated contour



Il precedente grafico indica le varie misurazioni “recenti” effettuate dal 1983 ad oggi dai ricercatori che si sono impegnati in questi studi. Le tabelle che compongono il grafico partono dalla soglia di udibilità (in basso a destra) fino ad arrivare ai 100 *phons* (in alto a sinistra). Dai dati riportati nel grafico, Suzuki ed il suo team hanno creato un nuovo e più accurato set di curve isofoniche:



Dal grafico si possono notare alcune cose: data la mancanza di dati sperimentali per valori di intensità molto alti, la curva dei 90 *phons* non va oltre i 4 *kHz* e la curva dei 100 *phons* si ferma ad 1 *kHz*. Inoltre, dato che valori raccolti per la curva dei 100 *phons* derivano da un solo istituto, è stato scelto di utilizzare una linea tratteggiata. Per lo stesso motivo è stata usata la linea tratteggiata anche per la curva dei 10 *phons*.



In questa figura vengono paragonate le curve isofoniche di Robinson e Dadson standardizzate nell'ISO 226 (a sinistra) e le più antiche di Fletcher e Munson (a destra) con le nuove curve stabilite dal professor Suzuki e dal suo team. Il grafico a sinistra mostra che, ad eccezione della curva della soglia di udibilità, le nuove curve si trovano distintamente al di sopra di quelle di Robinson e Dadson. La deviazione tra i due set di curve isofoniche è particolarmente evidente nella regione di frequenze sotto ad 1 *kHz* per un livello di loudness tra i 20 e gli 80 *phons*.

Dalla figura a destra, che invece paragona le nuove curve con quelle di Fletcher e Munson, si può notare una notevole somiglianza delle curve a 20 e 40 *phons*. Tuttavia, per livelli di loudness superiori ai 40 *phons*, la differenza è evidente per frequenze inferiori ad 1 *kHz*. Nonostante qualche differenza, è importante notare come le curve di Fletcher e Munson vengano di pari passo, per un ampio range di frequenze, con le nuove curve isofoniche.

Nella regione di frequenze tra 1 e 2 *kHz*, un piccolo picco della portata di pochi *dB* è visibile nelle nuove curve ma non in quelle classiche. Questo picco è stato costantemente osservato negli ultimi studi e sembra corrispondere ad un piccolo calo nella HRTF intorno a questo range di frequenze. HRTF sta per *Head-related transfer function* e descrive come il padiglione auricolare, il canale uditivo e il torso modificano l'intensità dei suoni di diversa frequenza che arrivano, ad ogni orecchio, da posizioni spaziali diverse. Molto probabilmente, il picco tra 1 e 2 *kHz* non è evidente nelle curve isofoniche classiche perché Fletcher e Munson non misurarono affatto livelli di uguale intensità tra 1 e 2 *kHz* mentre Robinson e Dadson misurarono livelli di uguale intensità in un solo punto all'interno di questa regione frequenziale. Di conseguenza, questo picco può essere trascurato.

Una nuova famiglia di curve isofoniche è nata da 12 recenti studi. Un'equazione è stata creata per esprimere la relazione di uguale intensità percepita tra toni puri a differenti frequenze. Questa procedura ha reso possibile la creazione di curve continue da un numero discreto di dati. Salvo nella zona vicino alla soglia di udibilità ed a livelli molto alti di pressione sonora, l'equazione dà un'ottima descrizione dei risultati degli esperimenti. In generale, le nuove curve mostrano una chiara differenza con quelle riportate da Robinson e Dadson (che costituirono il primo standard ISO 226), differenza particolarmente rilevante

sotto alla frequenza di 1 *kHz*. Le curve *classiche* di Fletcher e Munson, invece, mostrano una globale similarità con quelle nuove, anche se iniziano poi a differenziarsi sopra ai 40 *phons*.

Basandosi su questi risultati, il team di Suzuki propose il progetto alla ISO/TC 43 che lo accettò e, in base al lavoro svolto da questi ricercatori, stabilì i nuovi standard dell'ISO 226 nell'Agosto del 2003.

CAPITOLO 3

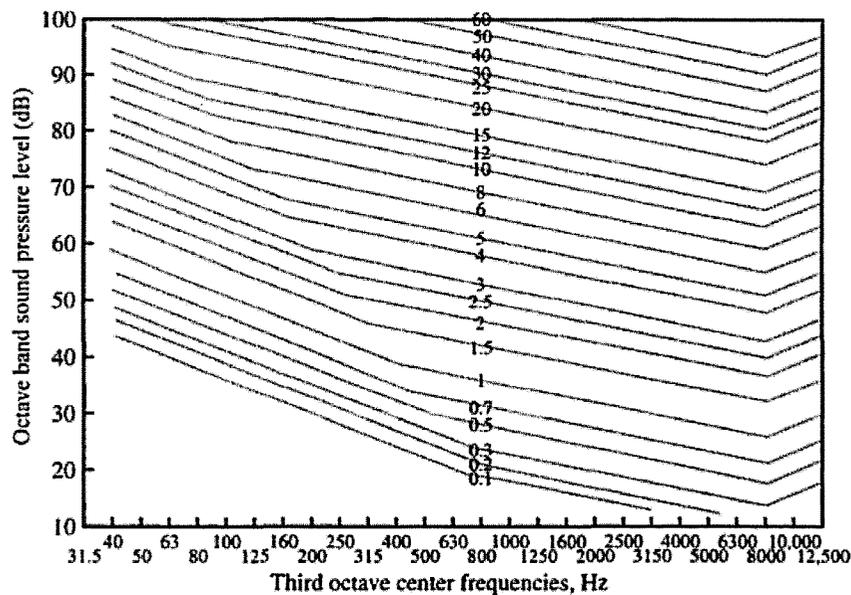
INTENSITA' SONORA PERCEPITA NEI SUONI COMPLESSI

3.1) CALCOLO DELLA LOUDNESS PER TONI COMPLESSI

Fino a questo momento, l'elaborato si è concentrato sullo studio e sulla misurazione della loudness di toni puri, caratterizzati cioè da una sola componente frequenziale. Nella quotidianità però, raramente abbiamo a che fare con toni puri. La quasi totalità dei suoni che raggiungono il nostro orecchio sono di natura complessa, caratterizzati cioè da diverse componenti frequenziali oltre alla fondamentale (*le armoniche*). Un tono complesso può essere scomposto in ogni sua componente frequenziale (*teorema di Fourier*) e la sua intensità è data dalla somma delle intensità di ogni sua singola componente. Nel calcolare la loudness di un tono complesso, verrebbe quindi istintivo calcolare l'intensità sonora percepita di ogni singola componente e sommarla con quella degli altri componenti frequenziali del tono complesso. Questo però non porta ai risultati sperati perché la sensazione sonora totale del tono complesso non dipende solamente dalle proprietà delle sue armoniche, bensì anche dall'interazione tra di esse. Se quindi un tono puro viene combinato con altri toni puri per formare un tono complesso, il suo contributo di loudness rispetto alla sensazione sonora totale sarà piuttosto basso, e sarà praticamente nullo nel caso gli altri componenti abbiano un'intensità maggiore e siano nella sua stessa regione frequenziale. Se, ad esempio, due segnali di pari intensità si sovrappongono all'interno di una banda critica, l'aumento di loudness è pari a 3 *phon*, mentre se i due segnali si trovano in bande critiche diverse si ha un raddoppio del livello in *son*, cioè un aumento di 10 *phon* per la loudness. Il fenomeno che entra in gioco è il *mascheramento*, già descritto nel capitolo 1.4.4.

Esistono comunque alcune tecniche per il calcolo dell'intensità percepita nei segnali complessi che tengono conto dell'effetto di mascheramento e che si basano sulla scala dei *son* descritta nel paragrafo 1.4.2. Nello standard OSI 532 sono definiti due metodi per il calcolo della loudness nei segnali complessi: il *metodo A* noto anche come *metodo di Stevens* ed il *metodo B*, anche noto come *metodo di Zwicker*.

Con il metodo di Stevens (OSI 532A), l'energia sonora di un suono complesso viene inizialmente divisa in ottave o terzi di ottava. Con l'utilizzo poi di un nomogramma di loudness (vedi figura sotto), viene assegnato un valore di loudness ad ogni banda di ottava. Utilizzando infine una sommatoria, si calcola la loudness totale partendo dai valori distinti di ogni singola banda. Questa tecnica tiene conto dell'effetto di mascheramento dando alla banda con il valore più alto un peso circa tre volte maggiore rispetto alle altre bande.



Analogamente a quanto avviene per il metodo di Stevens, anche con il metodo di Zwicker (OSI 532B) l'energia sonora del segnale complesso viene divisa in bande di terzi di ottava. Queste bande vengono poi convertite in una serie di bande critiche delle quali si calcola l'effetto di mascheramento. Infine, si calcola la loudness dalla somma delle bande.

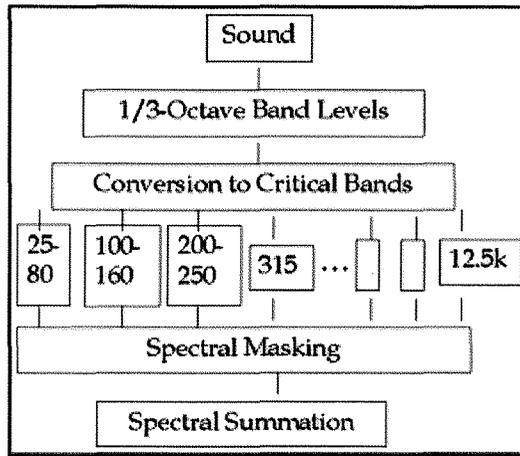


Diagram of an ISO 532B Computation

Questa tecnica richiede però una mole di calcoli molto elevata, viene quindi generalmente preferito il metodo di Stevens.

3.2) CONTROLLO DELLA LOUDNESS

I due metodi citati nel precedente paragrafo, sono stati per anni considerati punto di riferimento per il calcolo della loudness di toni complessi. Tuttavia, studi recenti, hanno dimostrato che queste tecniche si rivelano poco adatte per il calcolo di segnali dinamici tipicamente trattati nell'ambito della produzione musicale e del broadcasting.

Esistono molte applicazioni dove è necessario misurare e controllare la percezione della loudness. Per esempio nelle trasmissioni televisive e radiofoniche dove i contenuti cambiano frequentemente. In queste applicazioni l'audio dei contenuti cambia continuamente tra: musica, parlato e effetti sonori. La loro alternanza può creare delle differenze rilevanti nella percezione del loudness.

E' stato riconosciuto che la misurazione della loudness è necessaria nelle applicazioni broadcast, la normativa ITU (BS.1770) stabilisce come deve essere gestita la loudness in questo ambito. L'algoritmo raccomandato dall'ITU (BS.1770) permette di stimare il loudness calcolando l'energia media in frequenza pesata nel tempo, producendo un singolo valore. Il valore prodotto è il livello generale di loudness e viene calcolato in *Loudness Units (LU)*. Molti apparecchi di ultima generazione per il rilevamento di loudness (*loudness meter*) implementano questo algoritmo, considerato oggi il più affidabile.

3.2.1) Loudness nei programmi televisivi

Una questione ancora aperta, è la gestione loudness nei programmi televisivi delle varie emittenti. Il problema principale è la disuguaglianza di intensità sonora che viene percepita dall'ascoltatore quando cambia canale, specialmente nelle ore notturne, che lo obbliga ad un continuo aggiustamento del volume del televisore. L'esistenza di un preciso limite tecnico che stabilisce un tetto per i picchi del materiale audio consegnato per la messa in onda, non impedisce che i pubblicitari consegnino materiale particolarmente aggressivo e

disturbante, complici i sempre più potenti compressori multibanda disponibili sul mercato, che riescono ad aumentare il livello RMS mantenendo invariato il picco.

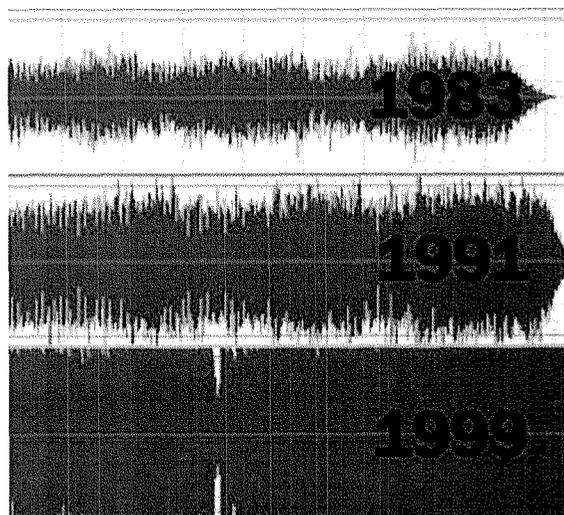
Il fatto di avere a disposizione un algoritmo che calcoli in maniera affidabile l'intensità sonora percepita di una trasmissione (come quello stabilito dall'ITU BS.1770), non è sufficiente. Pareggiando infatti la loudness media delle trasmissioni, senza quindi prenderne in considerazione la tipologia, il livello di un dialogo normale risulterebbe, nel caso di un presentatore nel concerto d'arpa, molto più alto di quello del film d'azione, essendo rispettivamente il dialogo più alto della media nel caso del concerto d'arpa, e più basso della media nel caso del film d'azione. E' perciò necessario identificare il livello del parlato ed utilizzarlo come riferimento per allineare i vari programmi. E' ovvio che l'estrema dinamica della voce non permette un'affidabilità assoluta, specie in programmi particolari o di breve durata: tra un bisbiglio ed un grido, specie considerando il tipo di inquadratura (primissimo piano, campo lungo), esiste una differenza di livello di alcune decine *dB*; tuttavia statisticamente il parlato, sul lungo termine e quindi certamente su un film, risulta abbastanza stabile ed equivalente a circa 70 *dB*.

3.2.2) *The Loudness War*

Il fenomeno della *Loudness War* (letteralmente guerra dei volumi) ha iniziato a prendere piede con la diffusione dei CD come supporto musicale. Alcuni casi risalgono però all'epoca dei vinili: molte compagnie di registrazione si misero a stampare compilation, e gli artisti ed i produttori le cui canzoni suonavano meno forte degli altri, pretesero che queste venissero ristampate per essere competitive. Tuttavia, i limiti di compressione del vinile, hanno fatto sì che l'incremento dei livelli di volume non raggiunse mai quello significativo che fu introdotto con l'avvento del Compact Disc e che raggiunse il culmine negli anni 2000. Il termine *Loudness War* si riferisce alla tendenza delle case discografiche, e dell'industria musicale in generale, a registrare, produrre e diffondere musica, anno dopo anno, con livelli di volume progressivamente sempre più alti, per creare un livello sonoro che superi, in intensità, i concorrenti e le registrazioni dell'anno

precedente. Un supporto come il CD ha però un'ampiezza massima di segnale limitata a 90 dB, questo significa che per avere un volume medio maggiore è necessario operare una compressione della dinamica. Comprimerne il range dinamico significa alzare il livello dei suoni di intensità minore ed abbassare (o meglio tagliare) i picchi di volume. Un uso eccessivo di questa tecnica può produrre dei *clips* e delle distorsioni in fase di registrazione compromettendo notevolmente la qualità della traccia. Questo avviene perché un media digitale non può emettere segnali più alti del suo *fondo scala* (ovvero il limite superiore che può essere effettivamente misurato) di 0 dBfs (*full scale*), in questo modo ogni volta che il picco di segnale oltrepassa questo punto, la forma d'onda viene tagliata (*clippata*). Dato che esiste un massimo livello di volume disponibile su una registrazione (all'opposto di quanto avviene nella riproduzione, in cui il volume è limitato dalla capacità degli amplificatori e dei diffusori), aumentando il volume complessivo di una canzone o di una traccia si crea un pezzo che uniformemente e complessivamente suona "forte" dall'inizio alla fine. L'intera canzone viene spinta contro il tetto massimo del volume, ciò crea una musica con una ridotta gamma dinamica (differenze minori tra le sezioni con un volume alto e quelle con un volume basso). In definitiva, la musica con una ridotta gamma dinamica perde la sua natura, è affaticante e insoddisfacente da ascoltare.

La seguente immagine mostra come sia notevolmente aumentato il volume di registrazione di un pezzo dei Metallica dal 1983 al 1999. Si può notare come il volume medio raddoppi ogni dieci anni circa:



Il motivo per cui avviene questa guerra dei volumi è fondamentalmente una questione commerciale. I dischi che hanno un volume maggiore vengono venduti maggiormente rispetto a quelli di volume minore questo perché la musica coi livelli sonori più alti è ascoltata e percepita più facilmente in ambienti rumorosi come una macchina, un treno o le strade trafficate della città. Inoltre, un volume più alto dei livelli sonori, può "rendere" meglio su sistemi di riproduzione di bassa qualità come i formati audio molto compressi, radio AM, televisori monofonici e telefoni.

Questo sta a dimostrare che l'interesse per la musica negli ultimi anni ha assunto un carattere sempre più quantitativo piuttosto che qualitativo. L'avvento del digitale ha portato sì ad un aumento generale della qualità della musica ma soprattutto ha dato la possibilità di poterla ascoltare in qualsiasi situazione ed in qualsiasi momento. Spesso ci troviamo ad ascoltare musica in maniera disinteressata, senza quindi prestarle attenzione. In una conversazione tra più persone, tendiamo ad alzare la voce nel momento in cui non viene prestata attenzione a quello che stiamo dicendo. Analogamente si comporta la musica moderna, creata in modo da farsi sentire anche quando nessuno sta ascoltando.

CONCLUSIONI

I risultati ottenuti dai test fatti in materia di intensità sonora percepita nei segnali audio, hanno dimostrato, in questa tesi, che l'argomento è più che mai aperto. La ricerca della perfezione, che è nella natura dell'uomo, ha spinto molti studiosi a cimentarsi in questo argomento, riprendendo studi compiuti in passato e, con l'aiuto del progresso tecnologico, riproponendone di nuovi e più accurati, cercando di dare una spiegazioni scientifica logica e chiara rispetto a quello che accade nella nostra testa quando riceviamo uno stimolo sonoro.

Come già detto più volte l'intensità sonora percepita è il risultato della combinazione di diversi fattori, sia di natura fisica che psicologica; questo significa che, pur conoscendo le leggi fisiche che governano il mondo della propagazione acustica, è essenziale proporre situazioni sonore studiate in laboratorio ad un vasto numero di ascoltatori selezionati (quindi con delle normali capacità uditive) per poter fare una stima di come esse interagiscano col nostro apparato uditivo.

L'elaborato sottolinea appunto l'importanza di riuscire ad ottenere il maggior numero di dati possibili da un esperimento: questo implica il fatto di dover ripetere lo stesso test, su un largo numero di ascoltatori, diverse volte, per poter ottenere una media precisa dei dati ottenuti. In una ricerca come quella presentata, il parere dell'ascoltatore è fondamentale.

Recenti studi hanno dimostrato come teorie piuttosto recenti e confermate (come le curve isofoniche di Robinson e Dadson del 1956), possano rivelarsi molto più inesatte di teorie formulate diversi anni prima (come le curve isofoniche di Fletcher e Munson del 1933). Come già detto, le curve isofoniche del 1956 sono andate a costituire lo standard internazionale OSI 226; se non ci fossero state le numerose ricerche svolte negli anni '80 e '90 ed il lavoro di Suzuki del 2001, lo standard si baserebbe ancora sulle curve isofoniche di Robinson e Dadson. Fu proprio questo ricercatore giapponese che con il suo team, grazie anche ai risultati ottenuti da vari team di ricerca di tutto il mondo nel corso degli

anni '80 e '90, dimostrò che lo standard fino ad allora in uso conteneva grossi errori e che i risultati degli studi compiuti nel 1933 avevano molta più somiglianza con i nuovi risultati rispetto a quelli del 1956. Questo vuole dimostrare come una ricerca continua ed un costante interesse verso le origini di questi studi, siano necessari per lo sviluppo di questa materia; è quindi necessario guardarsi indietro, analizzare e valutare quello che è stato fatto prima di noi e come è stato, per poi andare avanti a costruire sulle basi gettate dai pionieri degli studi sull'intensità sonora percepita.

Le scoperte fatte da questi ricercatori trovano oggi applicazione in diversi ambiti: dalla produzione musicale alla radiodiffusione, dall'edilizia alla medicina. Strumenti come ad esempio il *fonometro*, largamente utilizzati per misurare la pressione sonora, sono tarati in base alle *curve di ponderazione* che a loro volta derivano direttamente dalle curve isofoniche di Fletcher e Munson del 1933.

BIBLIOGRAFIA:

- Brian C. J. Moore, "An Introduction to the Psychology of Hearing", Elsevier Press, 5th ed.;
- D. W. Robinson et al., "A re-determination of the equal-loudness relations for pure tones", *Br. J. Appl. Phys.* 7 (1956), pp. 166–181;
- F. Müller, and E. Fichtl, "The measurement of equal loudness contours using a direct scaling procedure and validation by equal time contours," *Proc. Inter-noise* 94, 1069-1072 (1994);
- Glasberg Br. And Moore B.C.J., "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model" *J. Acoust. SOC. AM*: 125, 585 (2006);
- Harvey Fletcher and W.A. Munson, "Loudness, its definition, measurement and calculation", *J.A.S.A.* (1933);
- H. Moller, and J. Andresen, "Loudness at Pure Tones at Low and Infrasonic Frequencies," *J. Low Freq. Noise Vib.*, 3, 78-87 (1984);
- K. Betke,, and V. Mellert, "New measurements of equal-loudness level contours," *Proc. Inter-noise* 89, 793-796 (1989);
- Kingsbury, B. A. (1927). "A direct comparison of the loudness of pure tones," *Phys. Rev.* 29, 588–600;
- L. S. Whittle, S. J. Collins, and D. W. Robinson, "The audibility of low frequency sounds," *J. Sound Vib.* 21, 431-448 (1972);
- M. A. Bellmann, V. Mellert, C. Reckhardt, and H. Remmers, "Perception of sound and vibration at low frequencies," *Joint meeting of ASA and EAA: Forum*

Acusticum integrating German Acoustics DAGA Conference 1999, Berlin, Germany, (1999).

- Michael Talbot Smith, “Audio Engineer's Reference Book”, Focal Press, 2nd ed. (1999);
- R. Teranishi, “Study about measurement of loudness -On the problems of minimum audible sound-,” Researches of the Electrotechnical laboratory, No. 658, (1965);
- T. Poulsen, and L. Thogersen, “Hearing threshold and equal loudness level contours in a free sound field for pure tones from 1 *kHz* to 16 *kHz*,” Proc. Nordic Acoust. Meeting, 195-198 (1994);
- Yôiti Suzuki, et al., “Precise and Full-range Determination of Two-dimensional Equal Loudness Contours” (2003).