

La percezione uditiva

Massimo Grassi

3 giugno 2014

Indice

1	Cenni di acustica	3
1.1	Natura del suono	3
1.1.1	La velocità del suono	3
1.1.2	La lunghezza del suono	4
1.1.3	Trasmissione, riflessione, diffrazione	5
1.2	Per che cosa si caratterizza un suono?	6
1.2.1	Intensità e pressione	6
1.2.2	Frequenza e fase	8
1.2.3	Fourier, toni puri, toni complessi e rumori	9
1.2.4	Filtri	11
1.2.5	La rappresentazione grafica del suono: forma d'onda, spettro e spettrogramma	12
1.2.6	Modulazione d'ampiezza e modulazione di frequenza	13
2	Anatomia e fisiologia dell'orecchio	16
2.1	Orecchio esterno e medio	16
2.2	Coclea e membrana basilare	18
2.3	Dopo la coclea	18
2.4	Trasduzione	19
2.5	Musica, cervello e sordità	21
2.5.1	Musica e sordità	21
2.5.2	Musica e plasticità cerebrale	23
3	Intensità fisica ed intensità soggettiva	24
3.1	I confini dell'udito	24
3.1.1	Gamma dinamica dell'udito	24
3.1.2	Soglia assoluta	25
3.1.3	Soglia differenziale	27
3.1.4	Analisi dei profili spettrali	27
3.2	La funzione psicofisica dell'intensità soggettiva	28
3.2.1	Intensità soggettiva e frequenza	29
3.2.2	Intensità soggettiva e durata	30
3.2.3	Intensità soggettiva e inviluppo	31
4	Frequenza sonora e altezza tonale	32
4.1	Soglie e gamma d'esistenza	32
4.1.1	Soglia differenziale	34
4.2	La codifica dell'altezza tonale	34

4.2.1	La codifica dell'altezza tonale: toni puri	34
4.2.2	La codifica dell'altezza tonale: toni complessi	35
4.2.3	L'ipotesi di von Helmholtz	35
4.2.4	L'ipotesi di Terhardt e di Schouten	36
4.3	Altezza tonale dei toni musicali	39
4.4	Intervalli musicali armonici e melodici	40
4.5	Intervalli musicali armonici: la consonanza	41
4.5.1	La fusione tonale	44
4.6	Gli intervalli armonici	44

Capitolo 1

Cenni di acustica

1.1 Natura del suono

Il suono è una sequenza di variazioni di pressione che si sviluppa in un medium. Per quanto riguarda l'essere umano, il medium, nella maggioranza dei casi, è l'aria. L'esempio che segue ci aiuterà ad illustrare semplicemente che cosa succede quando si produce un suono. L'altoparlante del diffusore acustico del vostro impianto stereo trasforma un voltaggio elettrico in una variazione di pressione dell'aria. L'altoparlante si muove alternativamente avanti e indietro, “spingendo” e “tirando” l'aria. Questo movimento oscillatorio provoca una **condensazione** (l'alta pressione che si sviluppa quando l'altoparlante si muove verso l'esterno del diffusore) che si alterna ad una **rarefazione** (la bassa pressione che si sviluppa quando l'altoparlante si muove verso l'interno del diffusore). Tali variazioni di pressione vengono chiamate **onde sonore**. Le onde sonore sono onde longitudinali, ovvero, le particelle che trasmettono l'onda (le molecole nell'aria) oscillano avanti e indietro lungo la direzione di propagazione dell'onda, un po' come le spirali di una molla. Una volta generata, l'onda sonora si propaga in tutte le direzioni dello spazio. Questo è sempre vero, anche nel caso in cui la sorgente del suono sia di tipo direzionale così come può essere l'altoparlante di cui abbiamo parlato prima oppure la voce umana. In entrambi questi casi infatti, il suono si propaga per lo più frontalmente rispetto alla sorgente sonora. Tuttavia, come ben sappiamo, se la voce di un parlante è udibile in modo molto chiaro quando siamo di fronte al parlante, è vero anche che la stessa voce è udibile quando siamo alle spalle del parlante.

1.1.1 La velocità del suono

Il suono si propaga nel medium ad una certa velocità. Il valore di questa velocità dipende da numerosi fattori. Nell'aria il suono si propaga ad una velocità di circa 330 metri al secondo (m/s), 1188 chilometri all'ora, ovvero Mach 1. Ad un primo sguardo questa velocità può sembrare elevata, ma non lo è affatto se confrontata, ad esempio, con la velocità della luce che è di 299792458 m/s , ovvero circa trecentomila chilometri all'ora. Di fatto ci possiamo accorgere della “lentezza” del suono anche in situazioni quotidiane. Avete mai fatto caso che ai grandi concerti rock le ultime file del pubblico cantano in ritardo rispetto al cantante? Questo non succede perché i buontemponi usano ritrovarsi lontano

dal palco tutti assieme per cantare fuori tempo ma succede proprio a causa della limitata velocità di propagazione del suono nell'aria. Supponiamo di assistere ad uno di quei grandi concerti all'aperto, ad una distanza di 160 metri dal palco. Quando il cantante emette un suono, l'onda sonora giunge alle nostre orecchie con mezzo secondo di ritardo. Ecco che se decidessimo di cantare assieme al cantante ci troveremmo inevitabilmente a cantare in ritardo di mezzo secondo. Se mezzo secondo non vi sembra un tempo elevato sappiate che l'udito umano è in grado di apprezzare differenze temporali dell'ordine del microsecondo (μs) ovvero, differenze temporali molto ma molto più piccole di mezzo secondo.

La velocità del suono dipende da molti fattori. Ad esempio dal medium. In acqua il suono viaggia ad una velocità di circa 1484 Km/h . Se invece il medium è il ferro la velocità sale a 5120 Km/h (ed ecco spiegato perché l'indiano appoggia il suo orecchio sulla rotaia per capire in anticipo l'arrivo del treno). In generale, la velocità di propagazione del suono (c) in un medium è direttamente proporzionale alla rigidità del medium (C) ed inversamente proporzionale alla densità (ρ) del medium stesso secondo la seguente relazione:

$$c = \sqrt{\frac{C}{\rho}} \quad (1.1)$$

Ma la velocità di propagazione del suono dipende anche dalla temperatura:

$$c = 331.4 + 0.6T \quad (1.2)$$

dove la velocità c è espressa in metri al secondo e dove T è la temperatura dell'aria in gradi Celsius. Dalla equazione se ne ricava che con una temperatura di zero gradi la velocità di propagazione del suono nell'aria è di 331.4 m/s e che tale velocità aumenta fino a 343.4 m/s se la temperatura è di 20°C .

1.1.2 La lunghezza del suono

La velocità di propagazione del suono per sé non ha grossa importanza nella percezione uditiva (tralasciando ovviamente l'ambito dei grandi concerti all'aperto). Invece, è interessante conoscere la lunghezza d'onda di un'onda sonora. Il suono si propaga nello spazio. Ad esempio, se l'onda sonora che si propaga nello spazio è di tipo periodico (ovvero si ripete uguale a sé stessa dopo un certo tempo) possiamo conoscere la distanza che sarà necessaria all'onda per completare un ciclo completo. Tale distanza sarà uguale a:

$$L = cT = \frac{c}{f} \quad (1.3)$$

dove T è il periodo dell'onda sonora, ovvero il tempo necessario all'onda per completare un ciclo completo. Nell'equazione la velocità è espressa anche in funzione della frequenza del suono (f). La frequenza del suono è inversamente proporzionale al suo periodo:

$$f = \frac{1}{T} \quad (1.4)$$

ovvero, la **frequenza** del suono corrisponde al numero di cicli (o periodi) dell'onda sonora (ovvero, una singola alternanza di compressione e rarefazione) che avviene in un dato punto dello spazio e in un dato periodo di tempo. Le alte

frequenze sono comunemente associate ai toni acuti mentre quelle basse sono comunemente associate a quelli gravi. La frequenza di un suono è normalmente misurata in **Hertz** (Hz, 1 kHz = 1000 Hz).

A questo punto, possiamo calcolare la lunghezza d'onda di un suono in metri. Se il suono che stiamo suonando è di $20 - Hz$ (una nota poco più bassa della nota prodotta dal tasto più a sinistra di un pianoforte) lo spazio necessario all'onda per completare un ciclo completo sarà di circa 17 metri. Per contro, se la frequenza del suono fosse di $20 - kHz$ (ovvero, $20000 - Hz$) all'onda basterebbe una distanza di 17 millimetri per completare un suo ciclo. Tale marcata differenza ha delle ripercussioni sulla nostra vita di tutti i giorni. Ad esempio, le barriere acustiche che costruiamo per difenderci dal rumore del traffico sono più efficaci nel fermare le frequenze sonore alte (ovvero caratterizzate da una lunghezza d'onda corta) che le frequenze sonore basse frequenza che, in un certo qual modo, data la loro lunghezza "scavalcano" la barriera senza venire da questa attenuate in modo significativo.

1.1.3 Trasmissione, riflessione, diffrazione

Quando l'onda sonora si propaga nello spazio si possono verificare tre tipi di fenomeni: la riflessione, la trasmissione e la diffrazione. Se l'onda sonora incontra un ostacolo (ad esempio un muro) parte della sua energia verrà *trasmessa* all'ostacolo stesso che comincerà a vibrare. Quindi, se il suono d'origine era sufficientemente intenso, sarà possibile udire il suono anche al di là dell'ostacolo stesso. Questo è quello che succede quando riusciamo ad udire ciò che avviene nella stanza contigua a quella in cui stiamo.

Tuttavia, quando l'onda sonora tocca l'ostacolo, viene in parte *riflessa* dall'ostacolo stesso.

VEDI LUCIDI - PAG. 10

La capacità di riflettere il suono di un oggetto dipende in parte dalla sua forma geometrica e in parte dalla sua struttura. Ad esempio, la camere anecoiche sono camere costruite in modo tale da limitare il più possibile la riflessione del suono. Le camere anecoiche sono quelle camere le cui pareti sono coperte di lunghissimi prismi di vario materiale (ad esempio gommapiuma). Più artigianalmente, potrebbe esservi capitato di vedere un qualche amico ricoprire con le scatole delle uova le pareti del suo studio di registrazione. Ricordate però che camera anecoica non vuol dire camera silenziosa. Ricoprire le pareti di una stanza con dei prismi riduce la riflessione interna del suono ma non la sua trasmissione. Per avere una camera silenziosa (ovvero, una camera al cui interno viga il più totale silenzio così come in una cabina silente di quelle che si usano per le audiometrie negli ospedali) occorre equipaggiare la camera di muri pesantissimi, muri che, per quanto forte sia il suono che li colpisce, non trasmettano all'interno della camera la vibrazione sonora.

La *diffrazione* è l'ultimo dei fenomeni che caratterizzano la propagazione delle onde sonore. Le onde sonore possono aggirare gli ostacoli e per questo motivo riusciamo a sentire i suoni anche nel caso in cui tra le nostre orecchie e la sorgente sonora che stiamo ascoltando ci siano degli oggetti che impediscono il passaggio dell'onda sonora. Ma è possibile incontrare un altro esempio di diffrazione. Supponiamo di avere una radio accesa in una stanza dotata di una

piccola finestrella. Se fossimo fuori dalla stanza saremmo comunque in grado di udire la radio poiché, a partire dalla finestrella, l'onda sonora si rigenererebbe uguale a quella che era in origine. In altre parole, è come se la sorgente sonora (in questo caso la radio) si riproponesse tale e quale lì dove c'è l'apertura che permette il passaggio dell'onda sonora.

VEDI LUCIDI - PAG. 11

1.2 Per che cosa si caratterizza un suono?

Un suono si caratterizza per la sua frequenza (caratteristica che abbiamo in parte già trattato) per la sua ampiezza e per la sua fase. Inoltre, poiché i suoni avvengono nel tempo, il suono si caratterizza anche per una certa (durata anche se tale caratteristica non è propria dei suoni in quanto tali ma contraddistingue tutti gli eventi fisici). Ampiezza, frequenza, durata (e in parte fase) hanno una controparte soggettiva. Ad esempio, tanto più grande è l'ampiezza di un suono tanto più grande sarà l'intensità soggettiva che tale suono evoca. La frequenza sonora trova corrispondenza nella sensazione acuto/grave associata ai toni: tanto più alta la frequenza di un tono tanto più acuto questo ci apparirà. La durata trova la sua controparte nella durata soggettiva. La fase invece, non ha una controparte "diretta". Tuttavia, è facile dimostrare come l'essere umano sia sensibile alla fase. Infatti, la capacità di essere sensibili alla fase di un suono (o meglio, alla *differenza* di fase tra due suoni) è importante per la localizzazione delle sorgenti sonore nello spazio.

1.2.1 Intensità e pressione

Le onde sonore sono variazioni di pressione. Se il medium è l'aria le variazioni di pressione sono variazioni di pressione atmosferica. L'ampiezza dell'onda sonora è proporzionale al cambiamento di pressione che si ha durante una oscillazione della sorgente sonora. In sostanza, tanto più ampia è un'onda sonora tanto più grande la variazione di pressione che l'onda ha prodotto rispetto ad uno stato atmosferico di quiete, ovvero la *pressione atmosferica*¹. Tuttavia, le onde sonore si generano per una vibrazione. Quindi l'ampiezza dell'onda sonora sarà diversa di momento in momento. Durante la vibrazione ad esempio, ci saranno momenti in cui il cambiamento di pressione è nullo rispetto alla pressione atmosferica. Invece, in altri momenti il cambiamento sarà massimo. Per tale motivo, di solito, quando si parla di ampiezza di un suono ci si riferisce all'ampiezza media del suono o, in alternativa, all'ampiezza di picco. Però, se dovessimo calcolare l'ampiezza media di un suono ci troveremo di fronte ad un problema. Poiché la sorgente è in oscillazione, ad ogni momento di ampiezza positiva (i.e., pressione atmosferica maggiore di quella in quiete) corrisponderà un momento di pressione negativa di pari grandezza (i.e., pressione atmosferica minore di quella in quiete). In sostanza, la media aritmetica della pressione risulterebbe uguale a zero. È per questo motivo che si utilizza la cosiddetta *root mean square* (RMS). In italiano si usa l'espressione valore efficace. Ovvero, la media viene calcolata a partire sui valori di ampiezza elevati al quadrato (resi tutti positivi) e il risultato

¹La pressione atmosferica è la pressione esercitata dal peso dell'aria al di sopra del punto in cui la pressione viene misurata.

viene posto sotto radice quadrata. Per lo stesso motivo, quando ci si riferisce all'ampiezza di picco di un suono, si considera l'ampiezza in valore assoluto.

La pressione, e quindi anche la pressione sonora, sonora si misura in Pascal (Pa). In acustica, oltre a far riferimento alla pressione sonora si fa anche riferimento all'intensità sonora. L'intensità sonora è definita come l'energia sonora che passa attraverso un'unità di superficie in un secondo. La potenza a sua volta corrisponde all'energia trasmessa in un secondo. Poiché la potenza si misura in Watt, ne consegue che l'intensità sonora si misura in Watt per metro quadro (W/m^2). L'intensità di un'onda sonora è proporzionale al quadrato della pressione RMS dell'onda sonora stessa, ovvero:

$$I = P^2 \quad (1.5)$$

È bene ricordare che l'intensità di un suono (ed ovviamente la sua pressione) è in funzione della distanza: il suono prodotto da una sorgente sarà meno intenso se udito a, per esempio, venti metri di distanza piuttosto che a dieci metri di distanza. Esiste una specifica relazione che lega l'intensità sonora (o la pressione) e la distanza a cui questa viene misurata ed è la seguente:

$$I = \frac{1}{r^2} \text{ oppure } P = \frac{1}{r} \quad (1.6)$$

dove I è l'intensità del suono che stiamo misurando (oppure, dove P è la pressione del suono che stiamo misurando) ed r è la distanza da cui facciamo la misura.

I decibel

Sia l'intensità che la pressione sonora vengono comunemente espresse in **decibel** (dB), anche se, come abbiamo visto, le unità di misura di pressione ed intensità sono, rispettivamente, il Pascal e il Watt su metro quadro. La ragione di questo è semplice. Se andiamo a misurare la gamma di pressioni che interessa all'essere umano scopriremmo presto che questa copre valori che partono da 0.00002-Pa (la soglia assoluta per le intensità uditive) fino 20-Pa (pressione sonora alla quale oltre alla sensazione di suono si aggiunge quella di dolore). Insomma, la gamma di valori con cui si ha a che fare è decisamente molto ampia. Un modo per aggirare il problema è quello di trasformare le pressioni (o le intensità) in decibel. I decibel sono una unità di misura logaritmica che esprime la grandezza di una certa quantità fisica (di solito un'intensità) in rapporto ad una seconda quantità fisica di riferimento. In altre parole, poiché i decibel esprimono il *rapporto* tra due quantità e non la loro grandezza assoluta. Il calcolo della "distanza" in decibel tra due intensità avviene così:

$$\text{dB} = 10 \times \log_{10}(I/I_0) \quad (1.7)$$

dove, I è l'intensità sonora che vogliamo stimare in decibel e I_0 è l'intensità sonora di riferimento. Poiché l'intensità è uguale alla pressione elevata al quadrato, se volessimo calcolare i decibel a partire da una pressione sonora dovremmo modificare leggermente la formula come segue:

$$\text{dB} = 10 \times \log_{10}(P^2/P_0^2) = 20 \times \log_{10}(P/P_0) \quad (1.8)$$

dove P è la pressione e P_0 è la pressione di riferimento.

Come si vede dalla formula, la scala decibel è una scala logaritmica in base dieci. Di conseguenza, un incremento di 20 dB corrisponde ad un incremento di venti volte della pressione misurata rispetto alla pressione di riferimento. Invece, un raddoppio di pressione rispetto alla pressione di riferimento corrisponde ad un aumento di circa 6-dB.

Ora, vale la pena fare due considerazioni. La prima è di carattere generale. I decibel non sono una unità di misura esclusivamente acustica. Anche se comunemente quando si parla di decibel si fa riferimento all'intensità sonora, qualsiasi quantità fisica può, ovviamente, essere trasformata in decibel. La seconda considerazione è più interessante per chi si appresti a studiare la psicoacustica. Dire che un suono ha una pressione di 90-dB non significa nulla, poiché i decibel sono una scala relativa e non assoluta. Per capire se 90-dB sono tanti o pochi bisogna conoscere l'intensità di riferimento su cui quei 90-dB sono stati calcolati. Se l'intensità di riferimento fosse quella prodotta dal motore di un jet saremmo di fronte ad un suono di intensità elevatissima, perché 90-dB più intenso della intensità del suono prodotto dal motore di un jet (in altre parole, un suono in grado di romperci letteralmente i timpani). Se lo zero arbitrario fosse la soglia assoluta per l'intensità sonora di un pipistrello (un animale dotato di capacità uditive di gran lunga più elevate delle nostre) saremmo invece di fronte ad un suono di intensità per noi non particolarmente forte. È per evitare queste confusioni che, di norma, a fianco della scritta dB si trova una sigla (un acronimo) che specifica il valore dell'intensità di riferimento utilizzata per calcolare quella quantità di decibel. Per quanto riguarda la psicoacustica è stato scelto come intensità di riferimento un valore che, approssimativamente, corrisponde alla soglia assoluta per le intensità uditive dell'essere umano per un tono puro di frequenza 1000-Hz (ovvero, $2 \times 10^{-5} N/m^2$). I decibel che vengono misurati rispetto a questo zero si dicono dB SPL (Sound Pressure Level). Altri decibel che si utilizzano in psicoacustica sono i dB SL (Sensation Level) e i dB HL (Hearing Level). Ad ogni modo, i decibel verranno più accuratamente definiti nel capitolo dedicato all'intensità soggettiva. Inoltre, è proprio per i motivi appena descritti che spesso volte ci si trova di fronte a valori decibel negativi. Ad esempio, se leggiamo che "il soggetto era in grado di udire suoni di -5 dB SPL" vuol dire che quella persona è in grado di udire un suono la cui intensità è di 5 decibel più bassa della soglia media dell'essere umano per l'intensità sonora di un tono di 1000 Hz. Allo stesso modo (e quindi controintuitivamente) in certi apparecchi audio la manopola del volume è contornata da tutta una serie di valori decibel negativi che partono da $-\infty$ (quando la manopola del volume è chiusa) e arrivano a 0 (quando la manopola del volume è aperta) passando per tutta una serie di valori negativi di decibel. Com'è possibile che con il "volume a palla" faccio solo 0 dB? La spiegazione è la seguente. La reale intensità sonora prodotta dallo stereo dipende da moltissimi fattori (dalla distanza dalle casse, tanto per dirne uno). Il costruttore dello stereo però vi sta dando un'informazione e vi sta dicendo "guarda che se metti la manopola in quella posizione allora otterrai un suono la cui intensità è di tot decibel più bassa di quella massima ottenibile (ovvero 0)". Ed ecco spiegato il perché dei valori negativi.

1.2.2 Frequenza e fase

Come abbiamo detto in precedenza, la frequenza di un suono si misura in Hz, ovvero, in numero di cicli fatti dall'onda sonora nel tempo di un secondo. Spesso

si usa anche il kilohertz (kHz), ovvero quante migliaia di cicli ha fatto l'onda sonora nel tempo di un secondo. Tuttavia, quando si parla di suoni è più facile parlare di frequenze piuttosto che di frequenza. Esiste infatti un solo tipo di suono (il tono puro) che si caratterizza per un'unica frequenza. Più spesso (se non sempre, visto che i toni puri possono essere generati solo artificialmente) i suoni si caratterizzano per diverse frequenze contemporaneamente presenti. Ma approfondiremo questo tra poche righe.

L'ultima cosa che caratterizza un suono è la sua fase, ovvero, in quale momento il suono inizia. La fase di un suono di solito si esprime in gradi o radianti. Come riportato in precedenza, la fase, e più precisamente la *differenza* di fase, diventa importante per l'essere umano quando si tratta di localizzare le sorgenti sonore nello spazio, ed in particolare lungo l'asse orizzontale.

VEDI LUCIDI - PAG. 21

1.2.3 Fourier, toni puri, toni complessi e rumori

Fourier

I suoni del mondo difficilmente si caratterizzano per una sola frequenza. Di fatto, qualsiasi suono del mondo, può essere descritto ed ottenuto sommando un numero (per quanto grande) di singole componenti sinusoidali aventi diversa frequenza, ampiezza e fase. Ad esempio, se analizzassimo la voce umana che pronuncia una semplice vocale (e.g., "a"), scopriremmo che il suono è ricco, ricchissimo di varie frequenze, ognuna più o meno rappresentata (ovvero, ognuna con una certa, differente ampiezza) e ciascuna caratterizzata da una certa fase. Fourier è stato il matematico che per primo ha messo in luce questo fatto: la scomponibilità di un segnale, per quanto complesso, in una somma di elementi semplici sinusoidali. L'insieme delle componenti semplici alla base di un suono complesso (ovvero, attraverso la cui somma si ottiene un suono complesso) viene definita come lo *spettro* di quel suono. Quindi, se vogliamo riprodurre un suono, ad esempio un suono naturale, è sufficiente conoscere lo spettro del suono. A quel punto si genereranno un numero di sinusoidi pari a quelle contenute nello spettro (ovviamente rispettando anche la relativa ampiezza e fase) che si sommeranno poi tra loro. Il risultato della somma sarà il suono che volevamo riprodurre. Questo tipo di tecnica di generazione del suono viene chiamata sintesi additiva.

Nonostante il teorema di Fourier valga per tutti i segnali, è conveniente fare una separazione e descrivere separatamente i toni e i rumori. I toni si caratterizzano per spettri discreti, per contro i rumori si caratterizzano per il fatto di avere spettri continui.

Toni e rumori

Dal punto di vista acustico (ma in parte anche da quello percettivo) è possibile suddividere i suoni del mondo in due: i toni e i rumori. Come abbiamo detto, lo spettro dei primi è discreto, mentre lo spettro dei secondi è continuo. Ma c'è un'altra differenza che distingue i toni dai rumori. I toni si ripetono ciclicamente (ovvero periodicamente) uguali a sé stessi. I rumori invece si caratterizzano per variazioni di pressione che non hanno alcuna periodicità. Anche dal punto di vista percettivo i toni e i rumori si differenziano. I toni (ma non i rumori) si

caratterizzano per una certa altezza tonale. Tale caratteristica verrà descritta più avanti nel libro, nel capitolo dedicato alla percezione della frequenza sonora.

I toni

I toni puri Il suono più semplice che possiamo creare è un tono puro. Si dice puro quel tono caratterizzato per una configurazione di oscillazione che sia identica a quella di una sinusoidale. In pratica il “mattone” che va a costruire qualsiasi suono seguendo i principi del teorema di Fourier. I toni puri in natura praticamente non esistono. Due suoni naturali che somigliano al tono puro sono il suono del flauto e quello del fischio umano. Vista la sua semplicità strutturale, se conosciamo durata, frequenza, ampiezza e fase di un tono puro conosciamo praticamente tutto di quel tono.

VEDI LUCIDI - PAG. 28

I toni complessi Un tono il cui spettro contenga due o più componenti sinusoidali (e che quindi possa essere ottenuto sommando due o più toni puri) si definisce complesso. La voce umana è un esempio di tono complesso. Sono esempi di toni complessi, il fischio di molti uccelli, e il suono prodotto dagli strumenti musicali. Se si osserva lo spettro dei suoni complessi (ovvero l'insieme delle componenti sinusoidali che lo caratterizzano) si potrà vedere come un suono complesso si caratterizzi per un certo numero di componenti di frequenza differenti. Tuttavia, la distribuzione delle componenti in frequenza sarà discreta: prese le due componenti più vicine (ovvero più simili in frequenza) non ce ne sarà alcun'altra nel mezzo.

La quasi totalità dei suoni complessi che ci circondano è di tipo armonico. Nei suoni complessi armonici le varie componenti di frequenza di cui il tono è composto hanno una struttura particolare. Supponiamo che la prima frequenza (i.e., la frequenza più bassa) del tono complesso armonico sia 100-Hz. Se il tono è armonico, allora la seconda componente avrà una frequenza di 200-Hz (i.e., 100×2), la terza componente di frequenza avrà una frequenza di 300-Hz (i.e., 100×3) e così via. Quindi, in generale, lo spettro di un suono armonico complesso sarà così fatto:

$$f_n = f_0 \times n (\text{con } n = 1, 2, 3, \dots) \quad (1.9)$$

dove f_n è la frequenza della componente sinusoidale ennesima (detta anche armonica o parziale) e f_0 è la frequenza della componente più bassa, detta anche *frequenza fondamentale* o prima armonica. In altre parole, le parziali di frequenza di un suono complesso armonico sono ad una distanza (in frequenza) che è uguale alla frequenza fondamentale della serie armonica da cui si originano. Non è detto che un suono complesso armonico contenga tutte le armoniche della serie. Anzitutto non è detto che contenga un numero *infinito* di armoniche. In secondo luogo non è detto che tutte le armoniche siano presenti. Potrebbe infatti mancare una o più delle armoniche del suono. Due esempi di suoni complessi armonici “famosi” (perché usati ad esempio nella musica rock ai primordi della introduzione dei sintetizzatori) sono l'onda a dente di sega e l'onda quadrata. L'onda a dente di sega si ottiene sommando tutte le armoniche (la prima, la seconda, la terza, ...). Le armoniche devono tutte avere la stessa fase ma

l'ampiezza di ciascuna deve essere la metà rispetto a quella precedente. Così se la prima armonica (i.e., la fondamentale) ha una ampiezza di 1, la seconda avrà una ampiezza di 0.5, la terza di 0.25, e così via. L'onda quadra invece si ottiene sommando tra loro solo le armoniche dispari (la prima, la terza, la quinta, ecc.). Anche in questo caso le armoniche devono tutte avere identica fase. Tuttavia, l'ampiezza di ciascuna armonica è uguale al reciproco del numero armonico. Ovvero, la prima armonica avrà ampiezza 1, la terza ampiezza $1/3$, la quinta ampiezza $1/5$ e così via. Per entrambi i tipi di onde, è necessario un numero di armoniche infinito per generare, rispettivamente, un'onda "perfettamente" triangolare ed un'onda "perfettamente" quadrata, ma è possibile ottenere delle buone approssimazioni di onda a dente di sega e onda quadra anche utilizzando un numero finito di componenti di frequenza.

È possibile ottenere un tono complesso anche sommando tra loro parziali di frequenza che non sono in relazione armonica tra loro, in questo caso si parla di *tono inarmonico*. In psicoacustica spesso sono stati costruiti appositamente suoni inarmonici per verificare alcuni modelli della codifica dell'altezza tonale. In generale, il suono di molti strumenti musicali, pur avendo una struttura decisamente armonica, contiene almeno una qualche parziale la cui frequenza differisce da quella attesa seguendo la serie armonica.

I rumori Se i toni (siano essi semplici o complessi) si ripetono uguali a sé stessi ad ogni periodo, i rumori sono sempre diversi, ovvero non sono eventi periodici. Inoltre, se andiamo vedere il loro spettro, essi sono composti da un insieme continuo (e non discreto come nel caso dei toni complessi) di componenti di frequenza. Questo vuol dire che se il rumore contiene frequenze tra i 100-Hz e 200-Hz, date due qualsiasi componenti di frequenza che compongono il rumore (e.g., 150-Hz e 151-Hz) sarà possibile trovare infinite altre componenti che riempiono quello spazio di frequenza. Anche per i rumori esistono dei "rumori famosi". Il più famoso è forse il rumore bianco che, idealmente, contiene tutte le frequenze. In particolare, lo spettro del rumore bianco è piatto.

In psicoacustica si usano spesso i rumori. I rumori che si usano nella sperimentazione possono essere di vario tipo. Rumori a banda larga (ovvero simili ad un rumore bianco) oppure rumori a banda stretta, ovvero in cui siano rappresentate frequenze solo entro una certa banda. Oppure si utilizzano rumori passa-basso (che contengono solo le basse frequenze), passa-alto (che contengono solo le alte frequenze) o rumori banda-stoppa, che lasciano contengono tutte le frequenze eccetto quelle di una certa specifica banda di frequenza. Tutti questi rumori sono ottenuti filtrando un rumore di ampio spettro di frequenza (come ad esempio un rumore bianco) e rimuovendo le frequenze che si vogliono omettere.

1.2.4 Filtri

Come abbiamo appena visto per ottenere molti dei rumori che si usano in psicoacustica, si filtra un rumore bianco in maniera opportuna. Che cosa sia un filtro è presto detto. Il filtro di una sigaretta è uno strumento che trattiene alcune parti del fumo della sigaretta e contemporaneamente ne lascia passare alcune altre. I filtri che vengono utilizzati in psicoacustica hanno un funzionamento simile a quello dei filtri utilizzati per le sigarette: dato un segnale acustico in ingresso (ovviamente complesso) attenuano alcune delle sue parti di frequenza

lasciando inalterate delle altre. Ovviamente, i filtri possono essere utilizzati per filtrare qualsiasi segnale, quindi anche un tono.

Esistono diversi tipi di filtri e il nome che li caratterizza contraddistingue il tipo di azione che essi compiono sul segnale acustico che viene filtrato. I filtri possono essere di tipo **passa-alto**, se lasciano passare le frequenze alte, **passa-basso**, se lasciano passare quelle basse, **passa-banda**, se lasciano passare una certa gamma di frequenze, o **stoppa-banda**, se lasciano passare frequenze più alte e più basse della gamma di cui sopra. È importante accennare ad un altro tipo di filtraggio, quello che permette di ottenere un rumore con lacuna spettrale, poiché esso viene spesso utilizzato nella sperimentazione in psicoacustica. La lacuna spettrale si ottiene utilizzando contemporaneamente due filtri di tipo passa-banda le cui gamme di azione (in frequenza) non siano né contigue, né sovrapposte. Ad ogni modo, il rumore a lacuna spettrale può essere considerato un caso particolare di un rumore stoppa-banda.

È facile intuire quale può essere il risultato dell'azione di un filtro immaginando due situazioni che occorrono solitamente nella vita di tutti i giorni. La prima è il telefono. Il telefono non è un fedele riproduttore del segnale acustico. Invece, agisce come un filtro passa-banda e lascia passare solo le frequenze sonore comprese tra i 300 e i 3000 Hz. È per questo motivo che al telefono possiamo confondere la voce di una persona per quella di un'altra (cosa che molto più raramente succede nelle conversazioni "dal vivo"). È possibile inoltre avere un'idea quello che può essere l'effetto di un filtro di tipo passa-basso. Se ascoltiamo una voce umana attraverso un muro l'azione del muro nei confronti della voce è proprio di tipo passa basso. È per questo motivo che le voci attraverso i muri appaiono piuttosto cupe, poco brillanti, perché mancano di buona parte delle loro componenti sonore ad alta frequenza (che sono state attenuate dall'azione-filtro del muro). Lo stesso vale per qualsiasi suono sia udito "attraverso il muro".

Prima abbiamo visto che i suoni possono essere ottenuti per sintesi additiva. I filtri invece ci permettono di ottenere i suoni per sintesi sottrattiva. Supponiamo di voler riprodurre un suono complesso. Tale suono sarà caratterizzato da un certo spettro. Un modo semplice per riprodurlo è quello di partire da un suono che contenga tutto, e sottrarre (filtrando) tutte quelle parti di frequenza che non ci interessano. Insomma, un po' come la tecnica utilizzata dagli scultori che lavorano il marmo: si parte da un blocco di marmo e via via si tolgono tutte quelle parti che "coprono" la visione dell'oggetto finale che si vuole realizzare.

1.2.5 La rappresentazione grafica del suono: forma d'onda, spettro e spettrogramma

Un problema che incontra chi si avvicina allo studio del suono è che il suono non si vede. E quindi, bisogna rappresentarlo graficamente in qualche modo. Esistono tre modi di rappresentare graficamente un suono: la forma d'onda, lo spettro e lo spettrogramma. Nessuna delle tre rappresentazioni riesce a rendere evidenti, contemporaneamente, tutte le caratteristiche di un suono. Quindi, alternativamente, si usa una o l'altra a seconda di quale caratteristica del suono si voglia mettere in risalto.

La forma d'onda è una rappresentazione grafica di tipo bidimensionale e rappresenta l'ampiezza istantanea del suono in funzione della durata del suono stesso. Tale rappresentazione riesce a mettere bene in evidenza la fase del suono,

che può essere osservata guardando l'inizio della forma d'onda. Allo stesso modo è possibile cogliere immediatamente la durata di un suono ed il suo involuppo. È importante qui spiegare che cos'è l'involuppo di un suono. L'involuppo di un suono è quella linea immaginaria che congiunge tutti i picchi positivi di un'onda sonora. In sostanza, l'involuppo di un suono ha a che fare con la sua intensità poiché l'involuppo interpola quello che è l'andamento di intensità di un suono nel tempo.

VEDI LUCIDI - PAG. 60

In alcuni casi è possibile apprezzare anche la frequenza (o le frequenze) che caratterizza l'onda sonora. Ad esempio se l'onda è sinusoidale, la forma d'onda rappresenta appunto una sinusoidale e, calcolando la durata di un periodo, è possibile risalire alla frequenza del suono. Tuttavia, se il suono è complesso come un tono complesso oppure un rumore, difficilmente sarà possibile capire, guardando la forma d'onda, quali siano le frequenze che caratterizzano il suono che stiamo osservando.

VEDI LUCIDI - PAG. 30

Per questo motivo può essere utile ricorrere ad un diverso tipo di visualizzazione del suono, la visualizzazione di tipo spettro. Lo spettro di un suono rappresenta appunto lo spettro del suono, ovvero, il suo contenuto in frequenza. Come la forma d'onda si tratta di una rappresentazione bidimensionale. Quindi, nello spettro, viene rappresentata l'ampiezza delle componenti di frequenza di un suono in funzione del loro valore di frequenza. Grazie alla rappresentazione di tipo spettro quindi, è possibile apprezzare facilmente il contenuto di frequenza di un suono, tuttavia non è possibile, ad esempio, apprezzare la durata o l'involuppo né, tanto meno, la fase.

VEDI LUCIDI - PAG. 32

Il terzo modo per rappresentare graficamente un suono è lo spettrogramma. Differentemente dalle due rappresentazioni precedenti lo spettrogramma è una rappresentazione quasi-tridimensionale. La terza dimensione viene rappresentata con un colore. Lo spettrogramma rappresenta il contenuto in frequenza di un suono in funzione della durata. In aggiunta, il colore rappresenta l'ampiezza della componente di frequenza. Tanto più il colore tende al giallo tanto più quella gamma di frequenza è ampia. Tanto più il colore tende al blu tanto meno è ampia quella gamma di frequenza.

Poiché nessuna delle rappresentazioni riesce a cogliere tutte le caratteristiche del suono con solo disegno, spesso negli articoli di psicoacustica si affiancano due diversi tipi di rappresentazione, ad esempio, si affianca la rappresentazione a forma d'onda con la rappresentazione a spettro.

VEDI LUCIDI - PAG. 41

1.2.6 Modulazione d'ampiezza e modulazione di frequenza

I suoni di cui abbiamo parlato fino ad adesso sono suoni in qualche misura costanti. Ad esempio, sono suoni in cui l'ampiezza media non cambia per tutta la durata del suono, oppure dove il contenuto di frequenza del suono non

cambia per tutta la durata nel suono. Suoni costanti in ampiezza e contenuto di frequenza tuttavia sono abbastanza rari nel mondo reale. Invece, i suoni che udiamo normalmente sono cangianti nel tempo, sia in ampiezza che in contenuto di frequenza. Con modulazione d'ampiezza (AM) si intende una variazione dell'ampiezza media di un suono nel tempo. Con modulazione di frequenza (FM) si intende una variazione nel contenuto di frequenza di un suono nel tempo. Ovviamente le possibilità di modulare nel tempo sia l'ampiezza che il contenuto di frequenza di un suono sono innumerevoli. Di fatto, in psicoacustica, si preferisce utilizzare una modulazione di tipo sinusoidale. Le modulazioni sinusoidali, si caratterizzano per la profondità della modulazione (la grandezza dell'escursione in ampiezza oppure in frequenza) e per la frequenza della modulazione. Quando si modula un suono in ampiezza o frequenza il suono modulato viene chiamato portante mentre la funzione che modula viene chiamata modulante.

VEDI LUCIDI - PAG. 60

Per avere un'idea di come possa "suonare" una modulazione di ampiezza è sufficiente produrre una vocale costante in ampiezza e chiudere ritmicamente la bocca con la mano (un po' come fare il verso degli indiani per intenderci). Invece, un esempio di modulazione di frequenza è il suono di certe sirene in dotazione sulle automobili della polizia locale (quelle che suonano un po' come le sirene dei poliziotti dei telefilm americani per intenderci).

La modulazione di ampiezza pur modificando il contenuto di ampiezza di un suono (e non quello di frequenza) genera di fatto delle alterazioni nello spettro del suono. Supponiamo di avere una portante sinusoidale. Lo spettro di questa portante sarebbe semplicemente un unico punto di ampiezza pari alla ampiezza media della portante e di frequenza pari alla frequenza della portante. A questo punto moltiplichiamo la nostra portante per la modulante (la modulazione d'ampiezza di fatto è una moltiplicazione). Se andassimo a vedere lo spettro del suono modulato scopriremmo che alla componente di frequenza della portante si sono aggiunte altre due componenti di frequenza: le **bande spettrali laterali**. Se la frequenza della portante è pari a f_p e la frequenza della modulante è pari a f_m le bande spettrali laterali avranno frequenza pari a $f_p - f_m$, e $f_p + f_m$. Quando la profondità della modulazione di ampiezza è massima (i.e., 100%), ovvero quando l'involuppo del suono tocca lo zero, le bande spettrali laterali sono 6-dB meno intense della portante². Quindi, possiamo ottenere un suono modulato in ampiezza anche per sintesi additiva sommando tre componenti sinusoidali di frequenza f_p , $f_p - f_m$, e $f_p + f_m$.

VEDI LUCIDI - PAG. 57

È possibile ottenere una modulazione di ampiezza anche sommando due sole sinusoidi (invece che tre come nel caso precedente). La forma di modulazione che si ottiene è però leggermente diversa da quella sinusoidale. È il caso dei battimenti. Se si sommano tra loro due sinusoidi di frequenza differente si ottiene una modulazione d'ampiezza. La frequenza di questa modulazione sarà uguale alla differenza tra le frequenze delle due componenti di partenza. La profondità della modulazione invece sarà del 100% nel caso in cui le due sinusoidi

²Questo perché quando la modulazione è del 100% l'ampiezza della modulante è pari alla metà di quella della portante.

siano di pari ampiezza. Invece, se le ampiezze sono differenti, la profondità della modulazione sarà minore del 100%. Quando la differenza in frequenza tra le due sinusoidi è piccola (meno di 40-Hz), il battimento (i.e., la modulazione d'ampiezza) è chiaramente udibile. Se la differenza in frequenza è maggiore invece si percepisce una sorta di "rugosità".

Lo spettro che si ottiene modulando la frequenza di un suono è differente da quello che si ottiene con la modulazione di ampiezza. Poiché, grazie alla modulazione, la frequenza del suono varia di momento in momento, lo spettro quindi contiene un numero infinito di componenti di frequenza. Tuttavia, è possibile ottenere una modulazione di frequenza per sintesi additiva anche sommando un numero limitato di sinusoidi. Inoltre, se l'escursione in frequenza della modulazione è piccola, è possibile ottenere una modulazione di frequenza anche sommando semplicemente tre sinusoidi come nel caso della modulazione d'ampiezza.

Capitolo 2

Anatomia e fisiologia dell'orecchio

Lo studio della fisiologia del sistema uditivo è molto importante per capire alcuni (se non molti) dei fenomeni percettivi che caratterizzano la sensazione uditiva. Inoltre, chi conoscesse un po' della fisiologia che interessa alla percezione visiva, si accorgerà ben presto di una sostanziale differenza tra il sistema uditivo e quello visivo. L'elaborazione neurale dei segnali visivi avviene per buona parte a livello della corteccia. Per tale motivo, la ricerca sulla corteccia è molto sviluppata. Per contro, buona parte dell'elaborazione dei segnali uditivi avviene già a livello periferico, e più precisamente nella coclea, il più importante organo uditivo periferico. Ciò che arriva alla corteccia è in sostanza una informazione già fortemente "lavorata" ("lavorata" se la confrontiamo con il livello di elaborazione che subisce il segnale visivo prima di arrivare in corteccia). Se lo studio della corteccia visiva è molto sviluppato, lo studio della corteccia uditiva non lo è altrettanto e di fatto si sa ancora poco circa le operazioni che lì avvengono.

Il sistema uditivo periferico può essere suddiviso in tre parti: l'orecchio esterno, l'orecchio medio e l'orecchio interno. Una caratteristica degna di nota è che le tre parti si caratterizzano per medium diversi: il suono viene condotto in aria nell'orecchio esterno, viene trasmesso per via ossea a livello dell'orecchio medio (pur essendo l'orecchio medio "immerso" in aria), e viene trasmesso in medium liquido a livello dell'orecchio interno.

VEDI LUCIDI - PAG. 4

2.1 Orecchio esterno e medio

Il **padiglione auricolare** è la parte più esterna dell'orecchio esterno, ovvero, quella parte cartilaginea dove appoggiate gli occhiali. Il padiglione auricolare non è di fondamentale importanza per le nostre capacità uditive. Infatti, anche senza padiglione auricolare, siamo in grado di udire i suoni. I padiglioni auricolari sono più importanti in altre specie animali (pipistrelli, cani, ecc.) piuttosto che nell'essere umano dove sono troppo piccoli e rigidi (ovvero, non dotati della capacità di muoversi, ruotare, ecc.) per poter essere utili. Tuttavia, essi giocano un ruolo importante (anche se passivo) nella localizzazione delle sorgenti sonore

nello spazio poiché lì il suono che proviene dall'ambiente, per via della particolare forma del padiglione, incorre (in parte) in una particolare configurazione di riflessioni prima di entrare nel meato acustico.

Il meato uditivo (o dotto uditivo) è il canale che, a partire dal padiglione auricolare mette in collegamento il mondo esterno con la **membrana timpanica** (o timpano). Il meato uditivo ha una sua propria risonanza, così come la canna di un organo, e tale risonanza ci rende più sensibili alle frequenze sonore comprese tra i 1000 e i 6000 Hz. Questo è anche l'ultimo tratto nel sistema uditivo dove i suoni si trasmettono per mezzo dell'aria. In sostanza, il suono viene modificato dal passaggio nel meato un po' come quando la nostra voce si modifica quando parliamo attraverso il "tubo" della carta igienica. Ad ogni modo il suono percorre la lunghezza del meato uditivo per mettere in vibrazione una membrana che è posta proprio alla fine del meato: la membrana timpanica.

All'arrivo dell'onda sonora il timpano si mette in vibrazione. Le vibrazioni del timpano passano attraverso l'orecchio medio grazie a tre piccoli **ossicini**: l'incudine, il martello e la staffa, che sono le ossa più piccole dell'essere umano. In sostanza, se nell'orecchio esterno il suono viene trasmesso "nell'aria", nell'orecchio interno la vibrazione sonora, raccolta dal timpano, viene trasmessa in avanti, verso l'orecchio interno dal sistema degli ossicini. Il compito degli ossicini è di trasformare le variazioni di pressione di un ambiente pieno d'aria (il meato uditivo) in variazioni di pressione per un ambiente pieno di liquido (la **coclea**) nel modo più efficiente possibile. Tale trasformazione non è così semplice come può sembrare. Immaginate di "gridare" ad una persona che è sott'acqua. (Questo è in sostanza quello che succede quando il suono dall'orecchio esterno (medium aria) deve essere trasmesso, per mezzo degli ossicini, all'orecchio interno.) La maggior parte del suono da voi prodotto sarà riflessa dalla superficie dell'acqua stessa e pochissimo suono verrà trasmesso all'acqua. Il motivo della riflessione è che l'acqua ha un'**impedenza** molto più alta di quella dell'aria. Gli ossicini dell'orecchio medio risolvono questo stesso problema concentrando la vibrazione ricevuta dal timpano su di una superficie molto più piccola del timpano stesso, quella della finestra rotonda (o finestra ovale). Il rapporto tra la superficie del timpano e quella della finestra rotonda è infatti di 20:1. In sostanza, essi agiscono come una leva cosicché le vibrazioni ricevute dal timpano (grandi e deboli) vengono trasformate in vibrazioni (piccole e forti) sulla finestra rotonda. In termini elettrici gli ossicini possono essere considerati come un trasformatore d'impedenza.

C'è un altro fenomeno degno di nota che avviene a livello dell'orecchio medio. Attaccati all'incudine e alla staffa ci sono dei piccoli muscoli che si contraggono ad alte pressioni sonore (i.e., superiori ai 75-dB SPL), irrigidendo la risposta del sistema degli ossicini e riducendo in tal modo la grandezza delle vibrazioni trasmesse alla coclea. Questo meccanismo è molto efficace nel "tenere a bada" i suoni a bassa frequenza (i.e., inferiori ai 1000 Hz) ma non ci protegge dai suoni ad alta frequenza. Inoltre, questo meccanismo ha una certa latenza nell'attivarsi, dai 60 ai 120 ms. Quindi la sua efficacia è limitata ai suoni intensi ma non impulsivi (come ad esempio uno sparo) che non riescono ad essere attenuati.

2.2 Coclea e membrana basilare

La vibrazione trasmessa dagli ossicini si trasmette quindi all'orecchio interno ed in particolar modo alla coclea. Nel dettaglio, è la staffa che premendo sulla finestra ovale trasmette la vibrazione sonora alla coclea. La **coclea** è una specie di tubo sottile, pieno di liquido, avvolto in una sorta di forma a conchiglia. Negli esseri umani, se fosse allungata, la coclea sarebbe lunga circa 3-4 cm. La coclea è piena di una soluzione ionica. La coclea fa parte dello stesso sistema neurale dei **canali semicircolari** ovvero, del sistema neurale che controlla la sensazione di equilibrio.

La coclea è percorsa per tutta la sua lunghezza da due membrane: la membrana basilare e la membrana di Reissner. Quindi, se si potesse osservare una sezione trasversale della coclea si potrebbe notare tre compartimenti. Questi tre compartimenti vengono detti scale: la scala timpanica, la scala media e la scala vestibolare. La scala media è, delle tre, la più interessante poiché è lì che avviene la trasduzione del segnale sonoro da vibrazionale a elettrico. È lì infatti che si trova l'organo del Corti, l'organo che contiene le cellule ciliate responsabili della trasduzione. Alla base dell'organo del Corti c'è la membrana basilare. Sopra la membrana basilare stanno le cellule ciliate, nel dettaglio l'orecchio umano ha una fila di cellule cigliate interne e tre file di cellule cigliate esterne. Le cellule ciliate sono così chiamate perché alla loro sommità sono presenti delle ciglia (le stereociglia) che vanno ad inserirsi in una ulteriore membrana: la membrana tectoria. Le terminazioni delle cellule ciliate che fuoriescono dall'organo del Corti si raggruppano per formare il nervo acustico.

2.3 Dopo la coclea

I segnali elettrici in uscita dalla coclea vengono convogliati nel nervo vestibolo-cocleare, l'ottavo nervo cranico.

Essa vibra in risposta ai suoni che arrivano all'orecchio ma in modo particolare. Infatti, la membrana basilare può separare le diverse componenti di frequenza di un suono in modo meccanico. Ogni punto della membrana basilare è sensibile ad una ristretta gamma di frequenze. Alla base, verso la finestra rotonda, la membrana basilare è fine e rigida. Questa porzione è molto sensibile alle alte frequenze. Invece, la parte terminale della membrana basilare, detta apice, è relativamente spessa e flessibile ed è particolarmente sensibile alle basse frequenze. Le proprietà della membrana basilare variano in modo continuo tra i suoi due estremi. Cioè ogni punto della membrana basilare è massimamente sensibile ad una particolare frequenza sonora, detta anche **frequenza caratteristica**.

Tutto ciò è assolutamente cruciale per il nostro udito. Consideriamo cosa significa. Immaginate un contrabbasso e un ottavino [un flauto molto piccolo dal suono molto acuto] che suonino simultaneamente. La maggior parte dell'energia del contrabbasso si concentra nella regione delle basse frequenze. La maggior parte dell'energia dell'ottavino si concentra invece nella regione delle alte frequenze. I due strumenti stanno suonando contemporaneamente e le due onde sonore si mescolano nell'aria formando un'unica onda sonora, ovvero la somma delle due. Tuttavia, poiché le due onde coprono gamme di frequenza diverse, la membrana basilare può distinguere il suono che proviene da ciascuno dei due

strumenti cosicché udiamo chiaramente due suoni distinti. La raffinata scomposizione spettrale che viene effettuata dalla membrana basilare ci permette anche di identificare i due suoni nei termini delle loro caratteristiche spettrali. La sensibilità dell'essere umano per le frequenze va da circa 20 Hz fino a 20 kHz. Tuttavia, la maggior parte dell'informazione acustica per noi importante (per esempio, il parlato e la musica) viene veicolata da componenti di frequenza inferiori a 4 kHz.

Lungo la membrana basilare ci sono delle fila di cellule cigliate. Le cellule cigliate, così come il loro nome suggerisce, sono dotate di minuscole ciglia, o più correttamente, **stereociglia**, che si estendono dalla loro sommità. Nella coclea ci sono una fila di **cellule cigliate interne** e tre fila di **cellule cigliate esterne**. Si pensa che le cellule cigliate esterne cambino le proprietà meccaniche della membrana basilare. Esse migliorano la sensibilità dell'orecchio amplificando efficacemente le vibrazioni della membrana in risposta a livelli sonori bassi e medi (c.a. 0-80 dB SPL) ma solo per le frequenze sonore vicine alla frequenza caratteristica di ciascun punto della membrana basilare. (Si dibatte ancora se a questa amplificazione corrisponda un reale aumento di energia metabolica fornita al movimento della membrana basilare.) Così come migliora la sensibilità, questo processo migliora anche la capacità della membrana basilare di separare le componenti di frequenza di un suono. Un orecchio sano può udire separatamente componenti di frequenza che siano separate di circa il 15-20% in frequenza (Moore, 1997; Plomp, 1964a).

Le cellule cigliate esterne possono essere danneggiate facilmente dalla musica suonata ad alto volume. Se ascoltata ad alto volume, la musica può causare perdite di sensibilità permanenti così come perdite di selettività in frequenza. I danni alle cellule cigliate esterne sono la causa principale di perdita dell'udito. Le cellule cigliate esterne possono anche venire disattivate temporaneamente da farmaci quali l'aspirina.

2.4 Trasduzione

Le cellule cigliate interne sono responsabili della trasduzione delle vibrazioni della membrana basilare nell'attività elettrica del nervo uditivo. Le vibrazioni sulla membrana mettono in moto le ciglia della cellula e le piegano da parte a parte. Tale movimento causa un cambiamento elettrico (depolarizzazione) nella cellula. Le cellule cigliate interne fanno sinapsi sulle fibre del nervo uditivo. Ogni fibra nervosa è connessa con un particolare punto della coclea e quindi si caratterizza per una particolare frequenza caratteristica. Tale organizzazione **tonotopica** (data un'area di tessuto, le fibre nervose di cui è composta sono sensibili a frequenze differenti in una disposizione spaziale ordinata in modo progressivo) si mantiene fino alla corteccia uditiva.

Le fibre nervose uditive trasmettono informazione sotto forma di sequenza di impulsi elettrici (**potenziali d'azione**). In generale, maggiore è l'ampiezza del suono maggiore il numero di impulsi per secondo che viaggiano lungo l'assone della fibra nervosa. In aggiunta, per frequenze al di sotto di circa 5000 Hz, le fibre nervose uditive tendono a produrre impulsi che sono sincronizzati con il periodo della onda sonora entrante. Tale fenomeno viene chiamato **ancoraggio alla fase**, perché i neuroni tendono a produrre impulsi in tempi, o fasi, particolari, del ciclo dell'onda sonora (vedi Figura 4). L'ancoraggio alla

fase è il risultato delle proprietà delle cellule cigliate interne le quali si depolarizzano solo quando le ciglia sono piegate in una direzione (lontano dal centro della coclea). Benché i neuroni non producano normalmente impulsi a velocità maggiori di poche centinaia al secondo, l'informazione può essere combinata tra i vari neuroni in modo da riuscire a rappresentare la velocità di ripetizione dei toni ad alta frequenza [maggiore di 5 kHz]. Così, se un neurone produce impulsi in corrispondenza del primo, terzo, quinto ecc. ciclo dell'onda sonora entrante, un altro potrebbe produrre impulsi in corrispondenza del secondo, quarto, sesto ecc. in modo che le due velocità combinate siano uguali alla frequenza di ripetizione. Nella realtà i neuroni non si comportano in modo così regolare come questo esempio può lasciare intendere ma il principio è grosso modo lo stesso. Attualmente si pensa che l'ancoraggio alla fase possa essere utilizzato dai centri uditivi superiori per determinare l'altezza tonale dei toni.

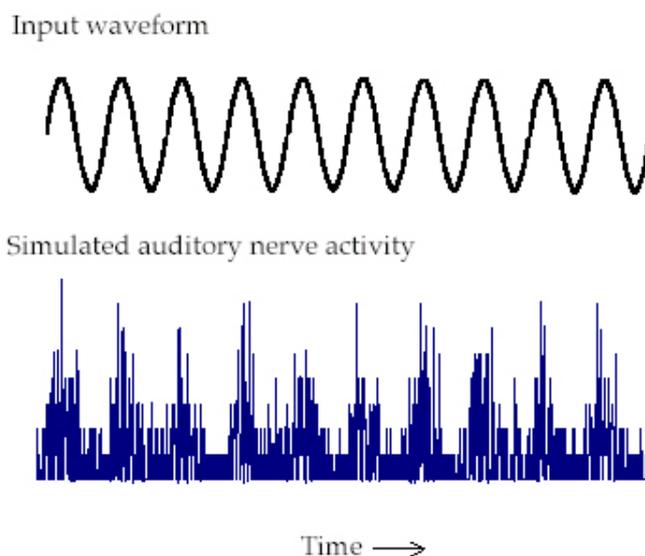


Figura 2.1: Una simulazione dell'attività del nervo uditivo in risposta ad un tono puro di 250 Hz. La simulazione rappresenta la somma dei potenziali d'azione di 500 fibre nervose. Le fibre nervose mostrano un chiaro ancoraggio alla fase, con un picco d'attività ad ogni ciclo della forma d'onda entrante.

Il nervo uditivo si connette al nucleo cocleare nel tronco dell'encefalo. Differentemente dal sistema visivo, nel quale la retina è connessa quasi direttamente (via talamo) alla corteccia visiva, nel sistema uditivo ci sono diversi nuclei subcorticali importanti (il nucleo cocleare, l'oliva superiore, il lemnisco laterale, il collicolo inferiore, il nucleo genicolato medio del talamo) che processano l'informazione neurale prima che questa sia convogliata alla corteccia uditiva che si trova sul lobo temporale.

Della corteccia temporale si sa ancora abbastanza poco anche se il numero di ricerche che ne investigano il funzionamento sta aumentando di anno in anno. Uno dei problemi che i ricercatori hanno incontrato nello studiare la corteccia

uditiva è che alcune tecniche di indagine (es la risonanza magnetica) mal si conciliano con la ricerca uditiva e musicale. Gli scanner che vengono utilizzati per la risonanza magnetica sono notoriamente rumorosi e conseguentemente è difficile investigare tutto ciò che coinvolge la presentazione di suoni. Tuttavia, per quanto concerne la corteccia uditiva, l'organizzazione rispecchia quella delle altre cortecce sensoriali (ad esempio quella visiva) con le aree primarie deputate alla elaborazione delle caratteristiche più semplici della stimolazione e le aree di ordine superiore deputate invece alla decodifica di caratteristiche più complesse.

2.5 Musica, cervello e sordità

Una delle caratteristiche più evidenti del training musicale, ovvero della prolungata attività di allenamento a cui si sottopongono solitamente i musicisti è che tale training evidenzia i suoi effetti sia sul comportamento del musicista che sulla sua risposta neurale ma anche nella anatomia delle strutture neurali stesse. Inoltre, poiché gli effetti del training musicale sull'udito sono noti, recentemente si è pensato di utilizzare i training musicali per migliorare l'udito nei casi di sordità.

2.5.1 Musica e sordità

Poiché il sistema uditivo si caratterizza sia per la presenza di una componente meccanica (es. gli ossicini), che per la presenza di una componente elettrica (es. le cellule cigliate) le sordità che possiamo incontrare sono riconducibili ad un malfunzionamento dell'una o dell'altra delle due componenti. In altre parole, la sordità può essere di tipo **trasmissivo** (dovuta ad esempio ad un malfunzionamento della catena degli ossicini) oppure di tipo **neurosensoriale** (dovuta ad esempio ad una mancanza di cellule ciliate). Inoltre, la sordità può essere congenita (ovvero essere presente sin dalla nascita) oppure progressiva, nel caso in cui la persona nasca udente ma diventi ipoudente nel corso del tempo. Un tipo di sordità che colpisce tutti in tal senso è la **presbiacusia** ovvero la perdita di sensibilità ai suoni acuti. Tutti, col passare degli anni, diventiamo via via meno sensibili ai suoni più acuti. La sordità viene solitamente classificata in più livelli: lieve, moderata, severa e profonda in funzione della quantità di udito residuo della persona. Le sordità di tipo trasmissivo raramente riescono ad essere di tipo severo o profondo al contrario di quelle neurosensoriali.

Le sordità severe e profonde possono essere trattate con l'impianto cocleare. L'impianto cocleare è sostanzialmente un orecchio bionico che bypassa l'azione delle cellule cigliate, che solitamente sono malfunzionanti e/o assenti nel caso dell'ipoacusia severa o profonda. L'impianto cocleare si compone di una parte esterna (simile nell'aspetto ad una normale protesi acustica) ed una parte interna inserita chirurgicamente che termina in un elettrodo che viene inserito a livello della finestra ovale e che viene fatto correre per quanto possibile lungo la coclea. L'elettrodo trasduce in impulsi elettrici ciò che la parte esterna dell'impianto riceve dall'ambiente.

L'impianto cocleare restituisce l'udito a persone che sentono poco o nulla. Tuttavia, l'udito che l'impianto cocleare offre è un udito impoverito se confrontato con quello biologico. Cosicché i portatori di impianto cocleare hanno delle evidenti lacune rispetto ai normoudenti. La prima lacuna è una peggiore capa-

capacità di comprendere il sesso del parlante a partire dalla sua voce (REF). Una seconda lacuna è una minore capacità di comprendere l'aspetto prosodico del linguaggio. Una terza lacuna è una peggiore capacità di percepire il parlato in situazioni rumorose, ad esempio in locali pubblici (un bar) dove siano presenti più parlanti contemporaneamente oppure dove ci sia un forte rumore di fondo.

Per ovviare a queste lacune negli ultimi anni si è pensato di offrire ai portatori di impianto cocleare dei training musicali. Infatti, è noto che i musicisti, rispetto ai non musicisti, sono più bravi nel cogliere le sfumature della voce (REF). In particolare i musicisti sono più abili (dei non musicisti) ad identificare la linea prosodica. Non da ultimo, i musicisti sono più abili nel percepire il parlato in situazioni rumorose (REF). Nonostante tali capacità percettive siano di solito il risultato di anni di training è lecito domandarsi se training musicali più brevi riescano comunque a migliorare le abilità uditive dei portatori di impianto cocleare. In aggiunta, molte delle persone che ricevono l'impianto cocleare si lamentano spesso del fatto che la musica, dopo l'intervento chirurgico, non suona più bella come prima. Gli impianti cocleari sono primariamente pensati, progettati e costruiti per rendere il parlato intelleggibile. Quindi, è normale che la riproduzione della musica sia scarsa o comunque non soddisfacente. In particolare, i portatori di impianto cocleare hanno una peggiore capacità di riconoscere i timbri sonori ed una peggiore capacità nell'identificare il contorno melodico delle frasi musicali, ovvero l'altalenarsi di acuto-grave che caratterizza le melodie. Anche in questo caso, è lecito chiedersi se un training musicale possa migliorare entrambe queste capacità restituendo, almeno parzialmente, la possibilità di godere la musica al portatore di impianto cocleare. E' importante sottolineare come il fatto di godere la musica non sia un piacere squisitamente percettivo. La musica è anche un evento sociale (la si ascolta assieme) e socialmente condiviso e condivisibile. Per questo motivo, ripristinare la capacità di apprezzare la musica può essere un ulteriore passo per ben-inserire il portatore di impianto cocleare in un contesto sociale.

Come annunciato precedentemente, solo recentemente si è valutata la possibilità di utilizzare i training musicali per migliorare l'udito (musicale e non) dei portatori di impianto cocleare. Per il momento i risultati rivelano come i training musicali siano in grado di migliorare l'udito musicale dei pazienti, sia per quanto riguarda la percezione dei timbri che per quanto riguarda la percezione del contorno melodico delle frasi musicali (REF). Tuttavia, per il momento, nessun ricercatore è ancora riuscito ad osservare se tali training trasferiscano i loro benefici anche su domini diversi da quello musicale come la percezione del parlato. Tuttavia, dato lo scarso numero degli studi condotti fino a questo momento, è prematuro giungere a qualsiasi tipo di conclusione.

I training musicali potrebbero anche essere utilizzati per migliorare l'udito delle persone anziane. I motivi sono esattamente gli stessi di quelli illustrati per i portatori di impianto cocleare. Nella ipoacusia che si verifica durante l'invecchiamento l'anziano perde abilità uditiva, intesa come la capacità di comprendere la linea prosodica del parlato o come la capacità di percepire il parlato in condizioni rumorose. Anche in questo caso, è lecito chiedersi se un training musicale possa in qualche misura migliorare l'abilità uditiva dell'anziano. I dati di letteratura dimostrano che sia i musicisti che i non musicisti perdono udito all'aumentare dell'età. Tuttavia, i musicisti sembrano preservare migliori capacità discriminative uditive rispetto ai non musicisti (REF). In futuro sarà interessante vedere se tali capacità trasferiscano alla percezione del parlato e

conseguentemente se i training musicali possano essere utilizzati per migliorare l'udito delle persone anziane.

2.5.2 Musica e plasticità cerebrale

L'intenso training a cui si sottopongono i musicisti prima di diventare tali ha degli effetti evidenti di plasticità cerebrale. In aggiunta, i training musicali solitamente iniziano molto presto. Cosicché i musicisti diventano tali dopo diversi anni di attività musicale magari iniziata ancora in tenera età. Inoltre, suonare uno strumento musicale è una attività complessa che richiede abilità percettiva (sia uditiva che visiva), unita ad abilità e controllo motorio di tipo fine. Tutte queste abilità devono infine essere coordinate. Non c'è da stupirsi quindi, se tale complesso sistema di attività produca degli effetti osservabili con le tecniche tipiche della psicobiologia.

Capitolo 3

Intensità fisica ed intensità soggettiva

L'intensità soggettiva (loudness in inglese) è il correlato percettivo dell'intensità (o pressione) sonora. Con le parole intensità soggettiva ci si riferisce quindi ad una grandezza psicologica del suono, differentemente dalle parole pressione o intensità si riferiscono a grandezze fisiche dell'ampiezza di un suono. I contenuti di questo capitolo sono divisi in due parti. Prima vengono esposti i risultati sperimentali di quegli esperimenti che hanno utilizzato le tecniche della psicofisica indiretta. Poi vengono presentati i risultati degli esperimenti che sono stati condotti utilizzando le tecniche della psicofisica diretta.

3.1 I confini dell'udito

3.1.1 Gamma dinamica dell'udito

La nostra capacità di udire intensità sonore ha due limiti: una soglia inferiore (più comunemente detta assoluta) ed una soglia superiore (detta anche soglia del dolore). Entro queste due soglie si può parlare di gamma dinamica dell'udito ovvero quella gamma di intensità sonore entro cui esiste la sensazione uditiva. La gamma dinamica dell'udito è ampia all'incirca 130 dB. Tuttavia, l'ampiezza effettiva di questa gamma dipende molto da quanto in là volete/riuscite a spingere la misurazione della soglia del dolore :-)

Tale ampia gamma dinamica ha rappresentato un problema per le teorie che spiegano come l'intensità venga codificata dal sistema uditivo. Infatti, se l'intensità fosse rappresentata solo dalla frequenza di scarica dei neuroni (così come suggerito dalla fisiologia) allora ci si dovrebbe aspettare una gamma dinamica molto più ristretta. Molte fibre nervose uditive non possono aumentare la loro frequenza di scarica all'aumentare dell'intensità quando il suono è superiore ai 60 dB SPL. Questo fenomeno viene chiamato saturazione. Tuttavia, ci sono delle fibre dotate di gamme dinamiche più ampie ed è possibile che tali fibre codifichino le intensità più alte.

3.1.2 Soglia assoluta

La misura della soglia assoluta umana è di solito fatta in due modi differenti. Il primo modo di misurare la soglia prevede che la misura venga fatta in una condizione dove all'ascoltatore sia permesso di usare entrambe le orecchie, e viene definita soglia MAF (Minimum Audible Field). Il secondo modo di misurare la soglia prevede invece che l'ascoltatore utilizzi un orecchio alla volta e che la misura venga effettuata per mezzo di una cuffia oppure con un auricolare. Tale soglia si chiama soglia MAP (Minimum Audible Pressure). Nella maggioranza dei casi le soglie uditive (sia MAF che MAP) vengono stimate utilizzando dei toni puri. Più raramente vengono usati dei segnali più complessi, ad esempio dei rumori a banda molto stretta.

Ecco come viene misurata una soglia MAF. L'ascoltatore viene posto all'interno di una stanza anecoica (o comunque caratterizzata da basse riflessioni) e i suoni che deve percepire vengono presentati per mezzo di casse acustiche. Le casse sono di solito posizionate ad 1 metro dall'ascoltatore e l'ascoltatore siede di fronte alla cassa che emette il tono. Si sceglie la frequenza sonora per cui misurare la soglia (e.g., 1000 Hz) e si presenta il tono. L'intensità del tono viene fatta variare per trovare quell'intensità in corrispondenza della quale l'ascoltatore riferisce di udire il tono il 50% delle volte. A questo punto l'ascoltatore viene fatto uscire dalla stanza e al suo posto viene messo un microfono (ovvero, si mette il microfono lì dove prima c'erano le orecchie del soggetto). Lo sperimentatore produce nuovamente il tono all'intensità (o pressione) sonora trovata in corrispondenza della soglia dell'ascoltatore. A questo punto, con uno strumento di misura collegato al microfono, si va a vedere quanti dB SPL corrispondono al tono presentato a soglia. Questa procedura viene ripetuta per tutte le frequenze sonore per cui si voglia misurare la soglia.

La misura della soglia MAP viene effettuata in modo analogo alla misura della soglia MAF con due importanti differenze: il tono viene veicolato per mezzo di cuffie (più spesso auricolari) che permettano di produrre il suono in prossimità del timpano, bypassando l'azione-filtro del dotto uditivo. Una volta trovato quel tono la cui intensità corrisponde alla soglia del soggetto, se ne misura l'intensità in dB SPL.

Se la misura della soglia MAF (o MAP) viene effettuata per un'ampia gamma di frequenze si osserva grosso modo questo risultato. La soglia (sia MAF che MAP) è bassa per frequenze medie e più alta per le frequenze sia alte che (in particolar modo) basse. La seconda cosa che può essere notata è che solitamente la soglia MAF è più bassa di quella MAP. La ragione risiede nel fatto che la MAF viene stimata in ascolto binaurale. Insomma, siamo più sensibili alle intensità sonore quanto i suoni stimolano entrambe le orecchie piuttosto che quando ne stimoliamo una per volta come nel caso della stima di soglia MAP. L'ultima cosa che si può notare è che la forma delle due curve non è esattamente identica. Infatti, spesso è possibile notare nelle soglie MAF (ma non nelle MAP) un più marcato abbassamento della soglia per frequenze attorno ai 2000 Hz. Il motivo è semplice. La misura di soglia MAF risente della risonanza del meato uditivo che amplifica proprio tale gamma di frequenza. Per contro, poiché la soglia MAP viene stimata bypassando l'azione del meato (e quindi escludendo l'azione di amplificazione dello stesso) l'azione di amplificazione del meato non può farsi sentire.

VEDI SLIDES - FIGURA 6

In alcuni studi è stata stimata anche la soglia superiore, ovvero la soglia del dolore. È facile immaginare come tali studi siano poco numerosi. Quindi le stime si fermano di solito al momento in cui la sensazione sonora comincia ad essere fastidiosa. Per suoni di 120 dB SPL più o meno, la normale sensazione uditiva si trasforma in una sensazione di fastidio. È interessante qui notare come questo avvenga indipendentemente dalla frequenza del suono che si sta udendo. In altre parole, se la soglia assoluta ha un profilo “U” la soglia del dolore ha un profilo relativamente “piatto”.

La soglia assoluta in audiologia

Se per qualche motivo vi capitasse di fare una audiometria in ospedale scoprireste che il grafico che rappresenta la vostra soglia non è il grafico a forma di “U” che avete appena visto, ma uno con profilo piatto. Insomma, l’audiogramma (così si chiama il grafico che rappresenta la soglia assoluta in funzione della frequenza sonora) in clinica è piatto. Il motivo è semplice. Come spesso capita le misure cliniche devono permettere una facile distinzione tra ciò che può essere una patologia e ciò che non lo è. Le audiometrie vengono fatte utilizzando delle cuffie e misurando la risposta di un orecchio per volta (così come le soglie MAP). Le cuffie sono calibrate (ovvero si conosce quanti dB SPL producono per un dato tono) e sono connesse ad uno strumento detto audiometro. Un grafico a forma di U non permetterebbe di certo la facile individuazione di una ipoacusia. Per contro, se il grafico fosse piatto (ovvero se l’audiogramma di una persona normoacustica fosse rappresentato da una linea piatta) ogni ipoacusia sarebbe facilmente identificabile poiché si manifesterebbe (graficamente) in una “deviazione dalla linea piatta”. In altre parole, in audiologia si preferisce rappresentare la soglia con un grafico piatto, rettificando ciò che altrimenti avrebbe una forma di U.

L’audiometria di norma viene fatta per un numero limitato di frequenze sonore. Per ciascuna di queste è stato stabilito un valore che permette di “rettificare” la curva a forma di “U” in un profilo piatto. Inoltre, per facilitare maggiormente la lettura delle ipoacusie, l’asse delle ordinate rappresenta i valori in ordine inverso, cosicché una ipoacusia si rappresenta con un avallamento (i.e., una deviazione verso il basso), piuttosto che con una gobba (i.e., una deviazione verso l’alto).

VEDI SLIDES - FIGURA 10

La tabella che segue riporta i valori che permettono la rettificazione dell’audiogramma. Una volta effettuata la conversione non si parla più di dB SPL ma di dB HL (Hearing Level) che sono i decibel in uso in audiologia.

VEDI SLIDES - FIGURA 11

I dB SL

Vale la pena qui di accennare l’ultimo tipo di decibel che si può incontrare in psicoacustica, i dB SL (Sensation Level). Immaginate di dover confrontare la performance di due gruppi di ascoltatori: uno normoacustico ed uno ipoacustico) in un compito uditivo. Poiché siete degli sperimentatori scrupolosi, volete che l’intensità soggettiva evocata dai suoni che presentate durante l’esperimento sia

la stessa, approssimativamente, per tutti i soggetti. Se presentaste suoni di pari dB SPL o pari dB HL però non otterreste questo risultato. Ad esempio, se presentaste i suoni a 65 dB SPL, tale intensità risulterebbe confortevole per il gruppo normoacusico ma (presumibilmente) troppo bassa per il gruppo ipoacusico. Un modo per ovviare a questo inconveniente è quello di misurare la soglia dell'ascoltatore che parteciperà all'esperimento, e di utilizzare come zero per il calcolo dei decibel proprio la pressione sonora in corrispondenza della sua soglia assoluta. Se fate questo state utilizzando i dB SL. Quindi, lo zero dei dB SL non è altro che la soglia dell'ascoltatore. Se vi capitasse di leggere un articolo dove gli autori scrivono che gli stimoli erano presentati ad un livello di 60 dB SL significherebbe che per ciascuno degli ascoltatori dell'esperimento è stata stimata la soglia assoluta e che successivamente per presentare gli stimoli si è amplificata l'intensità dei suoni fino a renderla 60 dB più elevata rispetto alla soglia del soggetto.

3.1.3 Soglia differenziale

La soglia differenziale per le intensità sonore viene di solito rappresentata in due modi: il rapporto di Weber (espresso in logaritmi) e ΔL :

$$\text{Rapporto di Weber (in dB)} = 10 \times \log_{10}(\Delta I/I) \quad (3.1)$$

$$\Delta L = 10 \times \log_{10}(I + \Delta I)/I \quad (3.2)$$

dove I è l'intensità di riferimento (o piedistallo) e ΔI è il più piccolo incremento (o decremento) che possiamo percepire di quell'intensità.

Per rumori a banda larga (uno stimolo che contiene un ampio spettro di componenti in frequenza) e per livelli superiori ai 20 dB SPL, il rapporto di Weber in funzione dell'intensità è approssimativamente costante (Miller, 1947), ed ha un valore di circa -10 dB (che corrisponde ad un ΔL di circa 0.4 dB). Questo significa che il rapporto del più piccolo incremento percepibile in intensità è proporzionale all'intensità del piedistallo. In sostanza, al di sopra di intensità di circa 20 dB si può dire che valga la **legge di Weber**, ovvero che la soglia differenziale è proporzionale alla grandezza fisica per cui la soglia viene misurata.

Per i toni puri si è scoperto che il rapporto di Weber diminuisce leggermente con all'aumentare dell'intensità fisica - in altre parole che la sensibilità aumenta all'aumentare dell'intensità - cosicché un grafico raffigurante $\log_{10}(\Delta I)$ in funzione di $\log_{10}I$ ha una pendenza di circa 0.9. Si pensa che questa approssimazione quasi perfetta alla legge di Weber ("near miss" in letteratura) rifletta la diffusione di eccitazione lungo la membrana basilare al crescere dell'intensità del tono. Ad intensità basse viene stimolata solo una piccola regione della membrana basilare (ovvero, quella regione "intonata" alla frequenza del tono puro), ma non appena l'intensità del tono cresce viene stimolata un'area più ampia della membrana. Questo surplus di informazione che viaggia nel nervo uditivo potrebbe migliorare la discriminazione per le alte intensità sonore.

3.1.4 Analisi dei profili spettrali

Benché, entro certi limiti, siamo sensibili all'intensità assoluta dei suoni, è la discriminazione delle intensità *relative* la più importante per identificare i suoni.

Ad esempio, l'identità della vocale /i/ non cambia al cambiare dell'intensità assoluta del suono. Sia che io gridi la vocale, sia che la sussurri, la vocale rimane per noi sempre il medesimo oggetto sonoro. L'identità della vocale /i/ è definita dal suo spettro. In altre parole, è definita dalla distribuzione in ampiezza delle sue componenti di frequenza.

Esiste una tipologia di esperimenti che investiga la capacità di individuare l'incremento in ampiezza di una singola componente di frequenza all'interno di uno spettro complesso, come può essere quello di un tono complesso. Tale classe di esperimenti viene chiamata analisi dei profili (profile analysis in inglese, ovvero analisi dei profili spettrali). Ecco come funzionano tali esperimenti. Si sceglie innanzitutto il profilo spettrale che si vuole investigare. Supponiamo ad esempio di scegliere un tono complesso armonico che contenga le armoniche dalla prima (la fondamentale) alla quinta. Per semplicità, supponiamo che l'ampiezza di tutte le armoniche sia pari. A questo punto si decide l'armonica bersaglio, ovvero quella armonica la cui ampiezza verrà variata per calcolare la sensibilità al particolare profilo spettrale del nostro tono complesso armonico. Supponiamo che l'armonica bersaglio sia la terza. Durante l'esperimento l'ascoltatore udirà due toni, un primo tono sarà il tono complesso armonico, il secondo tono sarà il tono complesso armonico in cui però l'intensità della sola terza armonica sarà aumentata di un certo valore ΔI . L'ascoltatore dovrà dire se i due toni sono uguali oppure diversi. L'ampiezza della armonica bersaglio verrà variata fino a che l'ascoltatore non sarà più in grado di dire se i due toni sono uguali oppure diversi, ovvero, quando l'ampiezza della componente bersaglio ha raggiunto il valore soglia. È importante sottolineare come in questi esperimenti l'ampiezza complessiva del tono venga resa causale entro una certa gamma di valori. Cosicché per rispondere, il soggetto deve necessariamente udire il profilo spettrale e non, ad esempio l'intensità complessiva del suono.

VEDI SLIDES - FIGURA 15

La nostra sensibilità alle intensità sonore è molto elevata se misurata con esperimenti di questo tipo. Ma c'è di più. La comparazione tra due forme spettrali può essere fatta in modo accurato anche su tempi lunghi, ovvero, se i due toni dell'esperimento sono presentati separati da un lungo (dell'ordine dei secondi) intervallo nel mezzo. La stessa cosa non accade quando stimiamo la semplice soglia differenziale per le intensità uditive. Se l'intervallo che separa le intensità che dobbiamo confrontare è lungo la nostra capacità di fare la comparazione decade immediatamente. In sostanza, la nostra buona memoria per i profili spettrali ci dà una misura della nostra (buona) capacità di discriminare (e ricordare) i timbri sonori.

3.2 La funzione psicofisica dell'intensità soggettiva

Ci sono tre tecniche principali per misurare l'intensità soggettiva: la **stima di grandezza**, la **produzione di grandezza** e l'**eguagliamento d'intensità**.

Nella stima di grandezza si chiede agli ascoltatori di fornire un numero che corrisponda all'intensità soggettiva di un dato suono, o di indicare quanto più intenso appare un suono rispetto ad un altro. Nella produzione di grandezza,

viene fornito un numero agli ascoltatori ed essi devono manipolare il livello di un suono fino a che la sua intensità soggettiva non eguaglia quel numero. In alternativa, gli si può presentare un suono standard e si può chiedere all'ascoltatore di manipolare il livello del secondo suono in modo tale che il rapporto tra le intensità soggettive dei due suoni corrisponda al numero fornito dallo sperimentatore. Nell'eguagliamento di intensità l'ascoltatore deve manipolare il livello di un suono fino a che l'intensità soggettiva di questo non appaia *identica* a quella di un suono di riferimento.

Queste tecniche (in particolare la stima della grandezza e la produzione della grandezza) possono essere utilizzate per stimare la funzione psicofisica dell'intensità sonora, ovvero, come vari la sensazione soggettiva di intensità al variare dell'intensità fisica. In generale si può dire che l'intensità fisica e quella percettiva sono fortemente correlate, ovvero, all'aumentare dell'intensità fisica aumenta anche l'intensità soggettiva. In aggiunta si può dire che la funzione psicofisica per le intensità sonore segua la legge potenza di Stevens, così come accade per molte quantità sensoriali. Per l'intensità soggettiva l'esponente corrisponde ad, approssimativamente, 0.2-0.3 per livelli superiori a 40 dB SPL cosicché:

$$L \approx kI^{0.3} \quad (3.3)$$

dove L è l'intensità soggettiva, I è l'intensità fisica e k è una costante. Per livelli inferiori a 40 dB SPL l'intensità soggettiva cresce rapidamente al crescere dell'intensità fisica. In generale, se moltiplichiamo per 10 l'intensità fisica percepiamo (approssimativamente) un raddoppio di intensità soggettiva

VEDI SLIDES - FIGURA 25

Stevens ha anche proposto delle unità di misura di intensità che fossero soggettive (e non quindi fisiche). Per quanto riguarda l'intensità sonora Stevens ha proposto il sone. Un sone viene definito (arbitrariamente) come l'intensità soggettiva evocata da un tono puro di 1 kHz presentato ad un livello di 40 dB SPL. Secondo Stevens, un suono di 2 sone è quindi un suono che, soggettivamente, evoca una intensità fisica doppia rispetto a quella evocata da un suono di 1 sone.

Se l'intensità soggettiva di un suono dipende per la maggior parte dalla sua intensità fisica è vero anche che l'intensità soggettiva può variare in funzione di almeno altri quattro fattori: la frequenza, la larghezza di banda, la durata ed in parte anche l'involuppo del suono.

3.2.1 Intensità soggettiva e frequenza

Per vedere tali effetti si utilizza solitamente l'eguagliamento di intensità, che permette di confrontare le intensità soggettive di due suoni aventi caratteristiche differenti. Ad esempio, è possibile confrontare le intensità soggettive di suoni aventi frequenza differente. Con tecniche simili è stato possibile ricavare i profili di equintensità ovvero, l'insieme dei suoni di frequenza differente che tuttavia evocano medesima intensità soggettiva (pur avendo intensità fisica differente).

VEDI SLIDES - FIGURA 25

Ecco come si ricava un profilo di equintensità. Si offrono al soggetto due toni puri: uno di 1000 Hz ed intensità nota (che funge da tono di riferimento) ed uno

di frequenza differente ed intensità manipolabile a piacere dall'ascoltatore. Dati i due toni, il compito dell'ascoltatore sarà quello di manipolare l'intensità del secondo tono in modo da renderla uguale all'intensità evocata dal tono di riferimento. Supponiamo che il nostro tono di riferimento sia di 30 dB SPL, ovvero un tono di bassa intensità. Supponiamo inoltre che il nostro tono manipolabile sia invece di 100 Hz. Se conducessimo un esperimento come quello appena descritto scopriremmo che l'ascoltatore (approssimativamente) deve aumentare l'intensità del tono manipolabile fino a più di 40 dB per ottenere due toni che gli appaiano di pari intensità.

In generale, si può dire che i toni di media frequenza sono quelli che ci appaiono più intensi mentre quelli a bassa e alta frequenza ci appaiono meno intensi dei toni di media frequenza. Stevens ha proposto anche in questo caso una unità di misura che riuscisse a cogliere come l'intensità soggettiva vari al variare della frequenza sonora. Tale unità di misura si chiama *phone*. Dato un tono (es. 100-Hz) la sua intensità è pari a n phon dove n è uguale all'intensità (in dB SPL) di un tono di 1000-Hz che evochi pari intensità soggettiva. In altre parole la scala phon è uguale ai dB SPL per un tono di 1000-Hz. Quindi, un tono di 1000 Hz presentato a 40 dB avrà una intensità soggettiva di 40 phone. Per contro, se il nostro tono ha una frequenza di 50 Hz la medesima intensità soggettiva (i.e., 40 phone) sarà ottenuta aumentando l'intensità fisica del tono fino a circa 65 dB.

Come abbiamo accennato in precedenza, i profili di equintensità rappresentano l'insieme dei suoni che evocano pari intensità soggettiva ma che hanno frequenza differente. La forma dei profili è marcatamente ad "U" quando l'intensità fisica del suono è bassa. Di fatto, questo è un fenomeno che avevamo già visto andando a studiare la soglia assoluta per le intensità uditive che non è altro che il profilo di equintensità a zero phone. I profili invece si appiattiscono all'aumentare dell'intensità fisica. Così per intensità sonore elevate (120 dB SPL) i profili sono per buona parte piatti. Per altro anche questo fatto ci era noto poiché la soglia superiore per le intensità uditive (la soglia del dolore) ha, come avevamo visto, un profilo piatto.

La frequenza altera l'intensità soggettiva di un suono anche in situazioni più complesse di quelle rappresentate nei profili di equintensità. In generale, dato un suono spettralmente complesso (ad esempio un rumore) la sua intensità soggettiva crescerà in funzione della larghezza di banda del rumore. Il caso è il seguente. Supponiamo di avere un rumore passabanda con frequenze da 750 a 1250 Hz e di intensità complessiva di 40 dB SPL. Se confrontassimo la sensazione di intensità evocata da questo rumore con quella evocata da un rumore passabanda con frequenze da 500 a 1500 Hz (ovvero con banda più larga) e pari intensità complessiva (ovvero 40 dB SPL) scopriremmo che il secondo rumore ci apparirà più intenso del primo.

VEDI SLIDES - FIGURA 37

3.2.2 Intensità soggettiva e durata

Un altro fattore che influenza la intensità soggettiva di un suono è la sua durata. Per quanto riguarda la soglia assoluta ad esempio, questa cresce al diminuire della durata del tono che si utilizza per stimarla. Tuttavia, se la durata del suono eccede i 200 ms la soglia rimane pressoché costante. Più in generale,

tanto più lungo è un suono tanto più intenso ci appare. Tale relazione però, così come per la soglia, vale entro una gamma di durate limitate cosicché se la durata dei suoni eccede i 200 ms circa intensità soggettiva non cresce più al crescere della durata.

VEDI SLIDES - FIGURA 39

È bene qui menzionare due ulteriori effetti che mettono in relazione la durata e l'intensità soggettiva. Tali effetti si chiamano adattamento all'intensità e fatica uditiva. L'adattamento all'intensità è quell'effetto per cui l'esposizione prolungata ad una certa intensità sonora fa sì che tale intensità ci appaia via via meno intensa. In tali esperimenti, l'ascoltatore stima la grandezza dell'intensità soggettiva mentre gli viene presentato un suono di lunga durata (e.g., un suono di qualche minuto). La stima viene fatta in diversi momenti della durata del suono. Ciò che si scopre è che la stima dell'intensità soggettiva diminuisce mano a mano che si ascolta il suono (pur essendo il suono costante in intensità fisica). Si parla quindi di adattamento all'intensità. Tale fenomeno tuttavia non è molto evidente per le intensità sonore così come accade in altre modalità sensoriali. In generale, l'adattamento all'intensità si verifica per lo più se il suono che si deve valutare è di bassa intensità (e.g., 30 dB SPL) e di alta frequenza. Per alte intensità e basse frequenze l'adattamento è minore.

La fatica uditiva è quel fenomeno per cui la soglia assoluta di un ascoltatore ha un innalzamento temporaneo dopo che l'ascoltatore è stato esposto per tempo prolungato ad un tono di forte intensità. In sostanza è quell'abbassamento temporaneo della nostra sensibilità uditiva che si verifica all'uscita dei concerti (ma non ha a che fare con quella sensazione di fischio che spesso si può percepire, quello si chiama tinnito o acufene). Gli esperimenti che misurano la fatica uditiva di solito si svolgono così: si misura la soglia dell'ascoltatore, poi lo si espone per un certo tempo ad un suono di intensità nota e quindi si rimisura la soglia. La seconda misura di soglia è solitamente più elevata della prima per via della fatica uditiva indotta dall'ascolto prolungato del tono.

3.2.3 Intensità soggettiva e involuppo

Infine, l'intensità soggettiva di un suono dipende in parte anche dal suo involuppo. Supponiamo di avere un suono con involuppo di tipo impulsivo, ovvero la cui intensità cresca improvvisamente nel tempo per poi calare progressivamente. Questo suono sembrerà meno intenso dello stesso suono se suonato al contrario nel tempo (ovvero, un suono caratterizzato da un progressivo aumento di ampiezza seguito da un improvviso decremento). La differenza in intensità soggettiva per i suoni che crescono o calano in intensità nel tempo è più marcata per i toni che per i rumori. Inoltre, non solo i suoni che crescono sembrano essere complessivamente più intensi dei suoni che calano in intensità nel tempo ma sembrano anche avere un maggiore cambiamento di intensità ed una maggiore durata soggettiva. Ancora una volta, entrambi questi fenomeni percettivi sono più evidenti se il suono è un tono piuttosto che un rumore, e, in parte, sono più evidenti quando la durata complessiva del tono è breve piuttosto che lunga.

Capitolo 4

Frequenza sonora e altezza tonale

L'altezza tonale è il correlato percettivo della frequenza sonora. In inglese esiste un termine specifico per riferirsi a questo particolare fenomeno percettivo: *pitch*. In italiano non esiste un termine specifico che possa essere associato al variare della frequenza sonora. Infatti nel linguaggio comune si parla indifferentemente di “altezza” sia quando ci si riferisce al fatto che un suono è più o meno intenso sia quando ci si riferisce al fatto che un suono è caratterizzato da una frequenza più o meno alta. In alternativa, in italiano si usano per parole “acuto” “grave” che meglio si avvicinano al concetto inglese di *pitch*. L'altezza tonale è quella caratteristica percettiva per cui siamo in grado di classificare i suoni come più o meno acuti (o più o meno gravi se preferite). L'American National Standards recita così: “Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus”.

Ma c'è di più, e il concetto di altezza tonale è, infatti, più esteso. Immaginate di premere un qualunque della parte centrale della tastiera del pianoforte. Esso produrrà una nota. Se ve lo chiedessi sareste in grado di riprodurre la stessa nota fischiando, oppure cantando. I tre suoni (il piano, il fischio, e il canto) sarebbero certamente differenti per timbro ma avrebbero in proprio la sensazione di nota. Quindi, con altezza tonale, non si intende esclusivamente la acutezza (o se preferite gravezza) dei suoni. Ma anche il fatto che alcuni suoni (i toni) siano in grado di indurre una particolare sensazione di nota dotata di una specifica altezza.

Questo capitolo è diviso in due parti. Nella prima vengono spiegati gli aspetti più squisitamente psicofisici legati alla percezione della frequenza sonora quali le soglie sensoriali. Nella seconda invece vengono spiegati i modelli che spiegano proprio quella sensazione di nota di cui si è appena accennato.

4.1 Soglie e gamma d'esistenza

L'orecchio è in assoluto molto sensibile alla ripetizione di configurazioni pressorie. Ad esempio, solo ed esclusivamente toni sono in grado di evocare una

sensazione di altezza tonale. Tuttavia, la nostra sensibilità alla ripetizione di configurazioni pressorie è possibile osservarla anche nell'ascolto dei rumori come nel caso del rumore periodico. Immaginate di generare un rumore della durata di qualche secondo. Se dopo averlo generato ne raddoppiaste la durata copiando ed incollando il rumore in coda a sé stesso (magari ripetendo l'operazione per un certo numero di volte) otterreste un rumore periodico. Il fenomeno singolare che si verifica all'ascolto di un rumore periodico è il seguente: si ode il rumore ripetersi uguale a sé stesso. Questo semplice fenomeno può dare una misura della sensibilità del nostro sistema percettivo al ripetersi degli eventi pressori.

Se per l'intensità sonora si può parlare con una certa precisione di soglie assolute (ad esempio quella inferiore) il discorso è un po' più complesso per la frequenza. L'essere umano è sensibile a frequenze sonore in una gamma che si estenda approssimativamente dai 30 Hz fino ai 16 kHz. Ma entrambi i limiti di questa gamma possono variare notevolmente. Ad esempio, la soglia superiore varia molto al variare dell'età e i bambini possono arrivare ad udire frequenze maggiori di 16 kHz mentre per gli anziani tale limite è di fatto troppo alto. Invecchiando infatti si perde la capacità di udire le alte frequenze. Per quanto riguarda il limite inferiore poi il discorso si complica ulteriormente. Anzitutto non è facile produrre suoni di così bassa frequenza. In secondo luogo, qualora li si riesca a produrre, siamo di fatto poco sensibili a tali frequenze sonore e per essere in grado di udire un suono di bassa frequenza siamo costretti a produrlo ad una intensità molto elevata. In ultima istanza, quando la frequenza sonora si abbassa la sensazione uditiva ad essa associata non si spegne così come accade per l'intensità ma, piuttosto, "cambia forma". Quanto la frequenza di un tono si abbassa al di sotto dei 30 Hz la sensazione tonale viene sostituita da una certa sensazione di tipo percussivo (un po' quello che si prova nelle discoteche dove le basse frequenze vengono presentate ad alte intensità sonore).

Inoltre, nella gamma di frequenze che spazia tra i 30 Hz e i 16 kHz siamo in particolar modo sensibili (ovvero, lì la sensazione di altezza tonale è più vivida) nella gamma tra i 30 Hz e i 5000 Hz. Tanto per avere un'idea, 30 Hz e 5000 Hz sono, circa, la nota più grave e la nota più acuta che può produrre un pianoforte, ovvero, la nota più grave e più acuta in uso in musica. Di fatto, se una melodia viene suonata utilizzando frequenze superiori ai 5000 Hz diventa meno riconoscibile (REF).

Per quanto riguarda la musica poi è importante sottolineare un altro fatto percettivo: le altezze tonali si ripetono ogni sette note. Se cominciate a suonare i tasti bianchi del pianoforte da sinistra verso destra scoprirete che la scala che stato suonando si ripete a diverse altezze. Questa ripetizione si verifica quando la nota che stiamo producendo ha una frequenza doppia rispetto alla nota da cui siamo partiti. Quando due toni sono in rapporto di frequenza $1/2$ si dice che sono in rapporto d'ottava, o, detto in altro modo, che l'intervallo tra le due note corrisponde ad un'ottava. In musica l'intervallo è la distanza (in frequenza) che intercorre tra due note. Tale distanza può essere di 1 (prima, es. do-do), 2 (seconda, es. do-re), 3 (terza, es. do-mi), 4 (quarta, do-fa), 5 (quinta, es. do-sol), 6 (sesta, es. do-la), 7 (settima, es. do-si) e 8 (ottava, es. do-do ottava superiore) a seconda di quante note si contano dal primo al secondo tono. Come dicevamo, se suoniamo una scala dopo l'arrivo all'ottava la scala sembrerà ripetersi uguale a sé stessa. In musica, la sensazione di altezza tonale si sdoppia. Da un lato abbiamo la sensazione di "acutezza" (*pitch height* in inglese) che aumenta all'aumentare della frequenza sonora. Dall'altro abbiamo

la sensazione di “nota” (*pitch* croma in inglese) che si ripete uguale a sé stessa ad ogni intervallo di ottava. Tale doppia sensazione è a volte definita come “teoria delle due componenti” (REF).

VEDI SLIDES - FIGURA 12

Tuttavia, la sensazione di “nota” è certamente più debole rispetto alla sensazione di acutezza. Infatti, se per suonare una melodia prendiamo a caso le note estraendole da diverse ottave la melodia non viene più riconosciuta (REF). La sensazione di acutezza e quella di nota attivano regioni differenti della corteccia uditiva (Warren, Upperkamp, Patterson, & Griffiths, 2003), ovvero, l’altezza tonale nel suo complesso verrebbe elaborata in più regioni della corteccia uditiva. Infine, la (forse più) famosa illusione uditiva (Shepard, 1964) illustra bene tale doppia proprietà della sensazione di altezza tonale. Ascoltando la scala di Shepard l’ascoltatore ode una scala musicale ascendente che però non ha mai fine, ovvero, si ode una scala che illusoriamente sale in altezza tonale salvo poi riprendere dal punto di partenza senza che l’ascoltatore se ne accorga.

4.1.1 Soglia differenziale

La soglia differenziale per le frequenze sonore è, in generale, più bassa per le frequenze al di sotto dei 5000 Hz e più alta per le frequenze al di sopra. Al di sotto dei 5000 Hz la soglia differenziale non varia in modo notevole al variare della frequenza (REF).

VEDI SLIDES - FIGURA 14

Invece, la soglia varia in modo notevole al variare dell’esperienza dell’ascoltatore. Ad esempio, i musicisti arrivano ad avere soglie differenziali fino a 10 volte più basse rispetto ai non musicisti (REF). Inoltre, la capacità di discriminare frequenze sonore, paradossalmente, può migliorare anche senza esercitarsi esplicitamente nel discriminare le frequenze (REF).

4.2 La codifica dell’altezza tonale

Esistono diversi modelli e spiegazioni che cercano di far luce su come venga codificata l’altezza tonale dei toni. Tali spiegazioni tuttavia, certamente si distinguono nello spiegare la sensazione di altezza tonale associata ai toni puri, e la sensazione di altezza tonale associata ai toni complessi.

4.2.1 La codifica dell’altezza tonale: toni puri

Ci sono due modi per rappresentare la frequenza di un tono puro nel sistema uditivo. In primo luogo, dato un tono puro di una certa frequenza, esso metterà in vibrazione (in modo massimo) uno specifico punto lungo la membrana basilare, mentre gli altri punti vibreranno in modo minore. Inoltre, se il tono ha frequenza inferiore ai 5000 Hz, la scarica delle fibre sarà sincrona rispetto ad una specifica fase del tono, ovvero, la scarica sarà, come abbiamo già visto, ancorata alla fase. Rispettivamente, tale due spiegazioni alternative vengono dette “rate-place coding” e “time-coding”.

VEDI SLIDES - FIGURA 19

Ci sono un certo numero di fatti che fanno pendere l'ago della bilancia verso una codifica di tipo time piuttosto che di tipo rate-place. In primo luogo, la sensazione di altezza tonale è più vivida per le frequenze al di sotto dei 5000 Hz. Inoltre, la soglia differenziale aumenta in modo considerevole per frequenze superiori ai 5000 Hz. In particolare, l'ipotesi rate-place sembra proprio che non riesca a spiegare proprio il valore della soglia differenziale per le frequenze sonore al di sotto dei 5000 Hz (REF). In sostanza sarebbe proprio l'ancoraggio alla fase (e conseguentemente la codifica di tipo time) ad essere la responsabile della sensazione di nota, specifica dell'altezza tonale. Mentre la più generica sensazione di acutezza potrebbe essere spiegata anche con la sola codifica di tipo rate-place. Per concludere, c'è un certo consenso tra gli autori nel ritenere che la codifica dell'altezza tonale dei toni puri sia di tipo time (o di tipo time combinato ai tipi rate-place) per frequenze al di sotto dei 5000 Hz. Mentre per frequenze al di sopra, venendo via via a mancare l'ancoraggio alla fase necessario alla codifica di tipo time, valga solo la codifica di tipo rate-place.

4.2.2 La codifica dell'altezza tonale: toni complessi

Se la codifica dell'altezza tonale dei toni puri è relativamente semplice, la codifica dell'altezza tonale dei toni complessi è relativamente più... complessa :-). Nei paragrafi che seguono si parlerà per lo più dell'altezza tonale evocata dai toni complessi armonici.

C'è innanzitutto un problema da risolvere. Un tono complesso è composto, almeno, da due componenti di frequenza. Potenzialmente quindi, nel caso di sole due parziali di frequenza, la sensazione di altezza tonale evocata dal tono potrebbe essere duplice, con una sensazione in qualche misura collegata alla componente di frequenza più bassa ed una collegata alla componente di frequenza più acuta. Ma se ci capitasse di ascoltare un tale tono scopriremmo ben presto che la sensazione è quella collegata alla frequenza più bassa ed in particolar modo alla frequenza fondamentale. Ovvero, dato un tono complesso armonico, la sensazione di altezza tonale che questo evoca è uguale a quella evocata da un tono puro la cui frequenza sia uguale alla frequenza fondamentale del tono complesso.

Esistono tre ipotesi circa la codifica di altezza tonale dei toni complessi. Una prima, ormai decisamente "storica", proposta addirittura da von Helmholtz (REF), una seconda ed una terza decisamente più recenti che invece, rispettivamente, chiameremo "modelli a riconoscimento di configurazione" e "modelli temporali".

4.2.3 L'ipotesi di von Helmholtz

La prima ipotesi circa la codifica di altezza tonale dei toni complessi risale a von Helmholtz ed è alquanto semplice (oltre che superata dai modelli più attuali). L'idea di von Helmholtz era semplicemente che la sensazione di altezza tonale corrispondesse alla frequenza fondamentale del tono complesso. In pratica, per dirla con parole contemporanee, dato un tono complesso armonico ci sarebbe un qualche detettore che individua la componente più bassa ed etichetta questa come la altezza tonale del tono complesso. L'ipotesi di per sé è semplice e di fatto

ragionevole. I toni nel mondo reale hanno sempre rappresentata la frequenza fondamentale. Ma in laboratorio è possibile sintetizzare toni complessi che non la contengano, ad esempio un tono che contenga solo la seconda, la terza, la quarta e la quinta armonica, ma non la prima. Che altezza tonale dovrebbe assumere un tale tono? Scondo l'ipotesi di von Helmholtz dovrebbe assumere quella della armonica più bassa, ovvero, in questo caso, della seconda armonica. Invece questo non succede e il tono complesso, di fatto, ha ancora una volta l'altezza tonale della fondamentale (che tuttavia in questo specifico esempio non è presente).

Ma l'ipotesi di von Helmholtz potrebbe reggere ancora. La frequenza fondamentale, assente nel suono, potrebbe reintrodursi nel sistema uditivo periferico grazie ai "toni di combinazione". I toni di combinazione sono il risultato di una distorsione che si verifica a livello del sistema uditivo periferico. In breve, la altezza tonale dei toni di combinazione è uguale alla differenza tra le frequenze dei toni che costituiscono il suono fisico. In altre parole, poiché nell'esempio di cui sopra la distanza in frequenza tra le armoniche è pari alla frequenza fondamentale (ad esempio la distanza tra la seconda e la terza, ma anche quella tra la terza e la quarta, ecc.) si potrebbe generare un tono di combinazione la cui altezza tonale sia pari proprio alla frequenza della fondamentale mancante. Tuttavia, più di qualche esperimento è riuscito a dimostrare come tale eventualità sia comunque da accantonare. Infatti, è possibile produrre un tono complesso senza fondamentale mancante e con l'aggiunta di un rumore a banda stretta proprio nella regione di frequenza della frequenza fondamentale. Lo scopo di tale rumore è proprio quello di mascherare l'udibilità di un eventuale tono di combinazione. A questo punto, se l'ipotesi di von Helmholtz fosse vera, l'altezza tonale del tono dovrebbe alzarsi e diventare pari alla seconda armonica, ma questo non succede e il tono ha sempre la altezza tonale della fondamentale. Inoltre, è possibile dimostrare che l'ipotesi di von Helmholtz sia falsa per mezzo di un tono complesso inarmonico. Immaginate di sintetizzare un tono con, ad esempio, tre componenti di frequenza spaziate di 200 Hz ciascuna ma che non facciano parte di una serie armonica con fondamentale a 200 Hz: quindi, 850 1050 1250 Hz. Dato un tale stimolo, e tenendo in mente l'ipotesi di von Helmholtz, potremmo prevedere che l'altezza tonale del tono sia di 200 Hz, poiché questo è l'altezza tonale del tono di combinazione che si sviluppa a partire da un tale tono inarmonico. Però questo non succede e in realtà l'altezza tonale percepita corrisponde a quella di un tono di circa 210 Hz.

4.2.4 L'ipotesi di Terhardt e di Schouten

Successivamente all'ipotesi di von Helmholtz, due ipotesi sono contese la spiegazione dell'altezza tonale dei toni complessi. Le due teorie di fatto sono contrapposte e differiscono nell'attribuire la sensazione di altezza tonale, rispettivamente, alle armoniche risolte (l'ipotesi di Terhardt) oppure alle armoniche irrisolte (l'ipotesi di Schouten).

Armoniche risolte ed irrisolte e regione dominante

Dato un tono complesso caratterizzato da molte armoniche parte di queste armoniche vengono dette risolte e parte invece vengono dette irrisolte. Grossolanamente si può dire che le armoniche risolte sono quelle con numero armonico

da 1 a 8 mentre quelle irrisolte sono quelle con numero armonico superiore a 8. Le armoniche risolte sono tali poiché il sistema uditivo periferico, a livello della membrana basilare, è in grado di “udirle separatamente”. In altre parole, queste armoniche riescono a produrre picchi discreti di oscillazione lungo la membrana basilare. Quindi, la capacità di selettività in frequenza del sistema uditivo periferico è in grado di udire separatamente queste componenti parziali di frequenza. Infatti, con un po’ di pratica anche dal punto di vista percettivo è possibile udire le prime parziali di frequenza di un tono complesso armonico, in particolare le prime cinque armoniche (REF).

VEDI SLIDES - FIGURA 35

La capacità di selettività in frequenza della membrana basilare invece non riesce ad “udirle separatamente” le armoniche irrisolte. Di fatto queste armoniche sono così vicine in frequenza che due o più parziali di frequenza vanno a cadere nell’area di uno stesso filtro uditivo. La conseguenza è che la risposta della membrana basilare a queste componenti di frequenza non è discreta come per le armoniche risolte. Il comportamento del sistema uditivo periferico può essere modellato con un banco di filtri passa banda. Quindi, se la risposta di un dato filtro uditivo per una componente risolta è la frequenza della componente stessa, per quanto riguarda le armoniche irrisolte questo non succede. Nel caso delle armoniche irrisolte la risposta del filtro non trasmette la frequenza di quelle parziali che cadono entro uno stesso filtro, tuttavia, rimane comunque qualcosa della periodicità del tono complesso. Supponiamo che le componenti irrisolte che cadono entro uno stesso filtro siano due. La periodicità di ciascuna componente verrà persa ma la periodicità della frequenza fondamentale del tono complesso verrà trasmessa. Le due parziali interagendo daranno luogo a battimento. Se le due parziali sono contigue, la frequenza del battimento sarà uguale alla frequenza della fondamentale del tono complesso. Ad esempio se la frequenza fondamentale del tono complesso è di 100 Hz e le due parziali sono la sedicesima (1600 Hz) e la diciassettesima (1700 Hz) la frequenza del battimento sarà di, appunto, 100 Hz. È bene ricordare tuttavia che “l’informazione di frequenza” trasmessa dal battimento sarà trasmessa dall’involuppo che si sviluppa nell’interazione delle due parziali. Il battimento è infatti la modulazione d’ampiezza che risulta dall’interazione di due parziali di frequenza.

Infine, è bene sottolineare un’ultima aspetto che caratterizza la percezione dei suoni complessi armonici: la regione dominante. Con regione dominante si intendono le prime armoniche del suono complesso. L’esistenza di una regione dominante è stata messa in luce da esperimenti dove si presentava un tono complesso con, di volta in volta, una delle armoniche leggermente stonata, ovvero con un valore di frequenza leggermente eccedente rispetto a quello previsto dalla serie armonica (es. 100, 200, 303, 400 and 500 Hz). In questi esperimenti viene offerto al soggetto un tono manipolabile in frequenza e il compito del soggetto è quello di eguagliare l’altezza tonale del tono complesso con manipolando quella del tono manipolabile fino a che i due non appaiano avere la medesima altezza tonale. Si è così scoperto che piccole alterazioni nel valore di frequenza delle prime armoniche sono in grado di cambiare l’altezza tonale del tono complesso. Differentemente da quello che si può immaginare, l’alterazione della frequenza della fondamentale non è la più efficace nell’indurre il cambiamento di altezza tonale nel tono complesso. Pare che invece siano rispettivamente la seconda e la

terza armonica quelle maggiormente in grado (se “stonate”) di alterare l’altezza tonale del tono complesso (REF).

VEDI SLIDES - FIGURA 33

Terhardt e Schouten

Secondo l’ipotesi di Terhardt l’altezza tonale di un tono complesso si genera a partire dalle armoniche risolte (REF). Le armoniche risolte sono “udite separatamente” grazie alla capacità di selettività in frequenza del sistema uditivo periferico. Secondo Terhardt, data una certa serie armonica, il sistema uditivo è in grado di riconoscere la distribuzione delle parziali di frequenza e quindi di estrapolare l’altezza tonale di un tono inferendo la frequenza della fondamentale a partire dalla distribuzione delle parziali. In altre parole, se il sistema uditivo riceve le frequenze 200, 300 e 400 allora la fondamentale è 100 Hz. Se invece nel tono sono presenti le frequenze 750, 1250 e 1500 Hz la fondamentale è 250. In pratica, secondo Terhardt, poiché la distanza tra due armoniche contigue è sempre pari alla fondamentale bastano due armoniche risolte per estrarre l’altezza tonale di un tono complesso, che, ovviamente può anche non essere dotato di frequenza fondamentale. In pratica secondo Terhardt esisterebbe una sorta di modello armonico che verrebbe messo in confronto con il suono che si sta udendo. Dal confronto verrebbe poi attribuita la frequenza fondamentale al tono complesso. Nota che il modello non è strettamente armonico e per questo motivo riesce a prevedere i piccoli cambi di altezza tonale che si hanno quando vengono alterati i valori di frequenza delle parziali della regione dominante (REF). Inoltre, il modello non specifica quale sia il meccanismo attraverso il quale le singole frequenze risolte vengono estratte. L’estrazione potrebbe essere fatta sia in forma rate-place che in forma time. Il vero limite dell’ipotesi di Terhardt è che non spiega come mai anche i toni composti di sole armoniche irrisolte siano in grado di evocare una certa sensazione di altezza tonale (anche se meno vivida). Inoltre, è possibile creare melodie anche

L’ipotesi avanzata da Schouten sostiene infatti che siano le armoniche irrisolte (e non quelle risolte) le responsabili della sensazione di altezza tonale dei toni complessi. Le armoniche irrisolte, interagendo creano battimento. Tale battimento ha una frequenza che è pari alla frequenza della fondamentale del tono complesso. La frequenza del battimento viene trasmessa per mezzo dell’involuppo che si genera unendo due parziali irrisolte di frequenza. Ma l’ipotesi di Schouten fallisce allo stesso modo di quella di Terhardt. Se le armoniche risolte non sono necessarie per la percezione dell’altezza tonale di un tono complesso non lo sono nemmeno quelle irrisolte. Un tono complesso dotato di sole armoniche risolte ha infatti una chiara altezza tonale. Inoltre, l’esistenza di una regione dominante per l’altezza tonale che coincide con le armoniche risolte mette in forte dubbio la possibilità che l’altezza tonale dei toni complessi possa essere estratta solo a partire da armoniche irrisolte.

VEDI SLIDES - FIGURA 39

Entrambe le soluzioni proposte da Terhardt e Schouten sembrano quindi fallire per il medesimo motivo: né le armoniche risolte né quelle irrisolte sono necessarie per la percezione dell’altezza tonale di un tono complesso. Negli anni

recenti sono infatti stati proposti modelli che utilizzano entrambe le armoniche per l'estrazione dell'altezza tonale. Tali modelli utilizzano il concetto di autocorrelazione. È bene ricordare come la difficoltà dell'estrazione dell'altezza tonale a partire da un tono complesso dipenda dal fatto che questo è caratterizzato da diverse componenti di frequenza. Immaginiamo la risposta del sistema uditivo periferico ad un tono complesso dotato di diverse armoniche sia risolte che irrisolte.

VEDI SLIDES - FIGURA 45

La risposta del sistema sarà ancorata alla fase per le armoniche risolte e, in un certo tempo, produrrà n spike in corrispondenza ad ogni fase della fondamentale, $2n$ in corrispondenza alle fasi della seconda armonica, $3n$ spike in corrispondenza della terza armonica e così via. Ma verrà estratta informazione di periodicità anche dalle armoniche irrisolte poiché qui osserveremo n spike ancorati alla fase dell'involuppo delle armoniche che cadono entro uno stesso filtro uditivo. L'idea dei modelli ad autocorrelazione è che il sistema confronti una tale configurazione con una rappresentazione della stessa ritardata per un certo intervallo nel tempo. Per ritardi pari al periodo della frequenza fondamentale il prodotto (ovvero, l'autocorrelazione) tra la configurazione e la rappresentazione sarà massimo, mentre lo stesso prodotto sarà necessariamente più basso per qualsiasi altro valore di ritardo.

VEDI SLIDES - FIGURA 46

4.3 Altezza tonale dei toni musicali

Fino a questo momento ci siamo occupati di capire come vari la sensazione di altezza tonale al variare della frequenza sonora. Complessivamente, all'aumentare della frequenza del tono corrisponde un aumento dell'altezza tonale percepita. Esiste però un'evidente anomalia in tale correlazione e si può osservare ascoltando le scale musicali. Se suoniamo ad uno ad uno i tasti bianchi del pianoforte a partire da quelli di sinistra per arrivare a quelli di destra si può notare come la sensazione di altezza tonale cresca mano a mano che ci spostiamo da sinistra a destra. Tuttavia, dopo un certo numero di note, le note della scala che stiamo suonando paiono ripetersi ad una altezza diversa. Le note musicali si ripetono effettivamente e non a caso assumono sempre gli stessi nomi (do, re, mi, fa, sol, la, si, *do, re, ...*). Questa ulteriore proprietà, caratteristica delle scale musicali, viene chiamata **altezza tonale cromatica** [pitch chroma] per distinguerla dalla altezza tonale pura e semplice [pitch height quando viene contrapposta a pitch chroma]. In alcuni testi si nomina tale duplicità come teoria delle due componenti. È importante però sottolineare un fatto. La relazione che lega le note che assumono lo stesso nome è debole. Ad esempio, data una certa melodia (ad esempio Fra Martino) non è possibile utilizzare, per costruire la melodia, delle note che hanno il nome giusto ma che provengono da ottave sbagliate.

Esiste in letteratura una famosa illusione uditiva (Shepard, 1964) illustra bene tale doppia proprietà della sensazione di altezza tonale. Ascoltando la scala di Shepard l'ascoltatore ode una scala musicale ascendente che però non ha mai fine, ovvero, si ode una scala che illusoriamente sale in altezza tonale salvo poi riprendere dal punto di partenza senza che l'ascoltatore se ne accorga.

L'illusione di Shepard è stata creata anche in forma di glissato, il famoso Glissato di Risset.

Come funziona l'illusione di Shepard? Nell'illusione si suona una scala cromatica, ovvero, la scala che otteniamo percorrendo tutte le note bianche e nere del pianoforte. Supponiamo di partire dalla nota do. Per ottenere l'illusione si suonano molti do, ad esempio tutti i do della tastiera del pianoforte. Tali note do vengono suonate contemporaneamente. Tuttavia, pur suonandole contemporaneamente, non si dà loro lo stesso risalto, ovvero la stessa intensità. Alcuni do verranno suonati intensamente, mentre altri do verranno suonati debolmente. La funzione che determina la intensità da attribuire a ciascun do è una funzione gaussiana centrata nel mezzo della tastiera del pianoforte in una particolare nota arbitrariamente scelta. Quindi, i do centrali saranno suonati intensamente, mentre i do acuti e gravi saranno suonati debolmente. Successivamente si suonerà la nota do (ovvero tutti i do \sharp) utilizzando anche qui lo stesso accorgimento: i do \sharp centrali verranno suonati debolmente e i do \sharp acuti e gravi verranno suonati debolmente. Quello che succede mano a mano che si passa da do a do \sharp a re e così via è che l'ascoltatore segue le note suonate nel registro centrale ma mano a mano che queste crescono in altezza tonale diventano via via sempre più deboli cosicché quando queste sono troppo deboli per poter essere udite distintamente la scala ricomincia e l'ascoltatore si ritrova ad udire la stessa identica nota che aveva udito per prima. Il glissato di Risset si basa sullo stesso principio ma il cambiamento di altezza tonale viene imposto in modo continuo e non discreto come nel caso della scala di Shepard. Un'ultima illusione (o paradosso) che utilizza i toni della scala di Shepard è il paradosso del tritono proposto originariamente da Diana Deutsch. In questo caso vengono suonate coppie di toni di Shepard che sono separati l'uno dall'altro di un intervallo musicale che corrisponde ad una quarta aumentata (detto anche tritono) ovvero l'intervallo che separa le note do e fa \sharp . Il tritono corrisponde a mezza ottava. I toni vengono presentati a coppie e il compito dell'ascoltatore è giudicare se la coppia descrive un intervallo ascendente oppure discendente. Poiché l'altezza tonale di entrambi i toni della coppia non è univoca, ogni coppia può essere interpretata sia in senso ascendente che discendente. Tuttavia esistono delle costanze nelle differenze individuali circa tali giudizi (vedi capitolo sull'orecchio assoluto).

4.4 Intervalli musicali armonici e melodici

Quando ascoltiamo la musica questa è composta da molte note, non da note isolate. Poiché la musica è un fenomeno complesso in psicologia spesso si tende a ridurre tale complessità a due sole note. Cosicché esiste una ampia letteratura che ha investigato la percezione di coppie di note siano esse presentate sincronicamente (ovvero intervalli armonici) piuttosto che diacronicamente (ovvero gli intervalli musicali melodici). In questa sezione vedremo come vengono percepiti gli uni e gli altri.

4.5 Intervalli musicali armonici: la consonanza

Se suoniamo la nota do e, successivamente, una dopo l'altra le altre note della scala musicale cromatica (ovvero, do#, re, re#, mi...) scopriremo presto che alcuni intervalli melodici suonano meglio di altri. Ad esempio, l'intervallo di ottava (do-do), quello di quinta (do-sol) e quello di quarta (do-fa) suonano meglio degli intervalli di seconda minore (do-do#), seconda maggiore (do-re) e quarta aumentata (ovvero il tritono, do-fa#). In pratica i primi intervalli vengono percepiti come consonanti mentre i secondi vengono definiti come dissonanti. Cosa vuol dire consonanti? Se leggiamo wikipedia scopriamo che In music, a consonance (Latin con-, with + sonare, to sound) is a harmony, chord, or interval considered stable (at rest), as opposed to a dissonance (Latin dis-, apart + sonare, to sound), which is considered unstable (or temporary, transitional). In more general usage, a consonance is a combination of notes that sound pleasant to most people when played at the same time; dissonance is a combination of notes that sound harsh or unpleasant to most people. This is a cultural concept as musics other than those from the western art music tradition, e.g. Balkan, Arabic, Chinese, do not follow this definition. Come si può vedere da questa definizione il concetto di consonanza è un soggettivo (l'intervallo consonante è piacevole) e statistico (per la maggioranza delle persone). Inoltre, apprendiamo che la consonanza ha anche una radice culturale poiché altre culture non sembrerebbero utilizzare questa bipartizione. In sostanza, la consonanza è quella caratteristica soggettiva per cui la maggioranza di noi giudica come piacevoli certi intervalli musicali e giudica come spiacevoli altri intervalli musicali, quelli dissonanti.

Esiste, tuttavia, una certa convergenza nei giudizi degli ascoltatori cosicché il fenomeno della consonanza si pone come degno di investigazione. Un'ulteriore aspetto interessante circa la consonanza è che se andiamo ad osservare il rapporto di frequenza che sussiste tra le note degli intervalli consonanti e lo paragoniamo con quello degli intervalli dissonanti scopriremo che i primi si caratterizzano per rapporti di frequenza tra numeri semplici. I secondi invece, sono caratterizzati da rapporti numerici più complessi. Ad esempio, nell'intervallo di ottava le note stanno in rapporto di frequenza 2/1 l'una con l'altra, nella quinta il rapporto è 3/2, nella quarta il rapporto è 4/3 e così via. Per contro, un intervallo dissonante come la quarta aumentata ha le note che sono in un rapporto di frequenza di 45/32, un rapporto numerico decisamente più complesso. La relazione tra la semplicità numerica che descrive l'intervallo e la eufonicità dell'intervallo stesso è nota fino dai tempi antichi e diversi autori hanno provato a darne spiegazione non da ultimi Pitagora e von Helmholtz.

In tempi più recenti sono state proposte due spiegazioni alternative del fenomeno della consonanza. La prima spiegazione sostiene che la consonanza venga appresa ascoltando i suoni del mondo. La seconda invece cerca una spiegazione nelle caratteristiche acustico/percettive degli intervalli che vengono giudicati come consonanti. In particolare, i sostenitori di questo secondo punto di vista hanno proposto due spiegazioni alternative che verranno illustrate in seguito.

Secondo alcuni autori (REF) noi giudichiamo come consonanti quegli intervalli musicali che sono maggiormente rappresentati nei suoni del mondo. In particolare, tali autori fanno appello alle caratteristiche acustiche della voce umana. Una analisi dettagliata della voce umana infatti rivela come lo spettro della voce stessa abbia molto rappresentati certi intervalli musicali (ad esempio

l'ottava) ed abbia poco o per nulla rappresentati altri intervalli (ad esempio la seconda minore, l'intervallo che intercorre tra do e do#). In sostanza, l'essere umano analizzerebbe lo spettro della voce (non è importante come tale analisi venga implementata) e poiché alcuni intervalli musicali sono più rappresentati di altri nella voce umana ecco che tali intervalli verrebbero giudicati come eufonici rispetto agli altri intervalli che sono poco o per nulla rappresentati. La conclusione che se ne trae è che la consonanza diventerebbe un fenomeno appreso. Questo tipo di spiegazione della consonanza non raccoglie tuttavia il massimo numero di consensi anche perché non è possibile validare sperimentalmente la spiegazione. Se è vero che i nostri giudizi di consonanza riflettono le caratteristiche acustiche dei suoni del mondo che ci sono più familiari possiamo trarre la conclusione che il concetto di consonanza sia condiviso universalmente. Invece, il concetto di consonanza è per lo più appannaggio della cultura musicale occidentale. Inoltre, un limite di questo tipo di spiegazioni è che non possono essere indagate sperimentalmente: non è possibile realizzare un esperimento che verifichi se tale ipotesi è (o meno) fondata.

Hanno invece raccolto più riscontro due ulteriori ipotesi circa la consonanza. Per entrambe il fenomeno della consonanza si origina proprio nelle caratteristiche dello stimolo musicale. Per per prima di queste ipotesi la consonanza è in funzione inversa della assenza di battimento dello stimolo musicale che stiamo ascoltando. Per la seconda la consonanza è in relazione diretta con la somiglianza armonica delle componenti di frequenza dei due toni che compongono l'intervallo consonante.

Secondo la prima delle due ipotesi, la sensazione di consonanza origina dalla assenza di battimento di certi intervalli e la dissonanza, per contro, origina dalla presenza di battimento di certi altri intervalli. Se consideriamo ad esempio un intervallo di ottava (es. do-do successivo) realizzato a partire da due toni armonici. Poiché le tutte le componenti di frequenza dei due toni coincidono, non c'è alcun tipo di battimento nell'intervallo e di conseguenza viene valutato come consonante. Plomp e Levelt (1965) prima e Kameoka e Kuriyagawa (1969) successivamente hanno dimostrato come i giudizi di consonanza fossero in funzione della assenza di battimento nello stimolo musicale. In particolare lo studio di Plomp e Levelt (1965) evidenziava come i giudizi di consonanza raccolti per intervalli formati da toni puri fosse in funzione della differenza in frequenza tra i due toni. In quella ricerca i due autori hanno raccolto dei rating di consonanza per diversi intervalli musicali composti di toni puri. Gli intervalli valutati dai partecipanti spaziavano da alcuni più piccoli di una seconda minore (do-do#) fino all'intervallo di ottava. Inoltre, gli autori hanno manipolato anche la frequenza del tono più basso che compone l'intervallo. I giudizi di consonanza raccolti erano molto coerenti ed erano pressoché gli stessi indipendentemente dalla frequenza del tono più grave che componeva l'intervallo. La consonanza era giudicata bassa per differenze di frequenze piccole ed alta per differenze di frequenza ampie.

VEDI SLIDES

Inoltre, la differenza in frequenza per cui i giudizi passavano da dissonante a consonante corrispondeva all'incirca al valore della banda critica. Se due toni cadono all'interno di una stessa banda critica non possono essere risolti dal sistema uditivo perché stimolano regioni troppo vicine lungo la membrana basilare e

la sensazione che ne risulta è appunto quella di battimento e rugosità. La banda critica vale approssimativamente un quarto di ottava. Plomp e Levelt (1965) ne dedussero che la consonanza fosse in funzione della presenza di battimento e rugosità nell'intervallo musicale. Una volta che tali fenomeni fossero assenti nello stimolo qualunque intervallo veniva giudicato come consonante.

Tuttavia, molti autori hanno criticato tale ipotesi. Innanzitutto, la spiegazione offerta da Plomp e Levelt (1965) includeva tra gli intervalli consonanti anche molti intervalli che storicamente erano sempre stati considerati dissonanti. Ad esempio, la settima minore (do-la \sharp) o la settima maggiore (do-si). Tali intervalli infatti eccedono abbondantemente la banda critica, non producono né battimento né rugosità e sono da considerarsi consonanti. Inoltre, esistevano già autori che offrivano spiegazione alternative (es. Terhardt, 1974). Secondo Terhardt, la consonanza poteva essere spiegata diversamente. Terhardt parte dalla considerazione che i toni musicali sono per lo più complessi e non semplici come quelli utilizzati da Plomp e Levelt (1965). Se confrontiamo le serie armoniche dei due toni complessi che compongono gli intervalli consonanti noteremo come le serie siano molto simili una con l'altra.

VEDI SLIDES

Secondo Terhardt (1974) era tale somiglianza la ragione della consonanza di certi intervalli: tanto più le serie armoniche che formano un intervallo musicale si sovrappongono (ovvero, ad esempio, hanno una sorta di denominatore comune che include tutte le armoniche della serie) tanto più l'intervallo viene percepito come consonante. Ad esempio, nell'intervallo di ottava c'è una identità completa tra le serie armoniche delle note che compongono l'intervallo. Per contro, nella settima minore non esiste alcuna armonica in comune.

Le ipotesi di Plomp e Levelt (1965) e di Terhardt (1973) sono state recentemente messe alla prova da uno studio di McDermott, Lehr e Oxenham (2010). In tale studio gli autori sono riusciti a investigare separatamente il ruolo della presenza di battimento e rugosità dal ruolo della somiglianza delle serie armoniche che compongono gli intervalli musicali. I risultati hanno dimostrato come i partecipanti preferissero stimoli senza battimento e ottenuti combinando serie armoniche simili. Tuttavia, solo la preferenza per le serie armoniche simili è risultata correlata con i giudizi di consonanza dati dagli ascoltatori. Inoltre, tale correlazione, anche se presente in tutti i soggetti, era tanto più forte quanti più anni il partecipante aveva studiato musica. In sintesi, questo studio dimostra come sia la somiglianza tra le serie armoniche a determinare la consonanza di un intervallo musicale. In aggiunta, sembra che studiare musica renda le persone ancora più sensibili rispetto a questo aspetto.

In sintesi, il fenomeno della consonanza sembrerebbe dipendere in parte dalle caratteristiche dei suoni che ascoltiamo (ovvero, il fatto che le serie armoniche dei toni che compongono l'intervallo siano simili) in parte da ciò che impariamo (la correlazione tra la somiglianza delle serie armoniche e il giudizio di consonanza è più accentuato nei musicisti). Tuttavia, è sempre importante sottolineare che la consonanza è un fenomeno soggettivo e che per questo deve essere investigata chiedendo dai giudizi ai partecipanti. In altre parole, data la intrinseca soggettività del fenomeno, è difficile che qualunque spiegazione sia soddisfacente a tutti i livelli.

4.5.1 La fusione tonale

Esiste, forse, una via alternativa per spiegare il fenomeno della consonanza. Secondo Stumpf esistono alcuni intervalli musicali armonici in cui le due note che compongono l'intervallo sono così simili tra loro che l'intervallo stesso viene percepito come fuso. La buona fusione tra le note che si verifica in tali intervalli è a volte così elevata che gli ascoltatori faticano a capire se l'intervallo è composto di una sola nota oppure di due note. È il caso dell'intervallo di ottava. Stumpf investigò sperimentalmente tale fenomeno e chiese a degli ascoltatori di giudicare il grado di fusione degli intervalli musicali. Con buona sorpresa l'autore notò che i giudizi di fusione correlavano molto con i giudizi di consonanza. In altre parole, secondo Stumpf, consonanza e fusione tonale sarebbero le due facce di una stessa medaglia: tanto più le note che compongono un intervallo sono ben fuse tanto più l'intervallo stesso viene percepito come consoante. DeWitt e Crowder (1987) hanno investigato sperimentalmente la fusione tonale descritta da Stumpf. Secondo i due autori, un ascoltatore che oda un intervallo musicale ci metterà più tempo per capire di quante note è composto l'intervallo quando questo è ben fuso rispetto a quando le note non sono ben fuse. Inoltre, è presumibile che l'ascoltatore faccia più errori ascoltando gli intervalli ben fusi piuttosto che gli intervalli mal fusi. Negli esperimenti di DeWitt e Crowder gli ascoltatori ascoltavano suoni che potevano essere intervalli musicali oppure note isolate. Il compito dell'ascoltatore era di dire il più velocemente possibile se il suono ascoltato era un tono isolato oppure un intervallo. Gli autori hanno osservato come i tempi di risposta fossero più lunghi per gli intervalli di ottava e quinta (considerati ben fusi e consonanti) e più brevi per gli intervalli di tritono e seconda (considerati mal fusi e dissonanti). Anche gli errori fatti dai soggetti rispecchiavano le attese degli autori. Tuttavia, gli esperimenti di DeWitt e Crowder (1987) non hanno permesso di ottenere un chiaro ranking per tutti gli intervalli musicali. Inoltre, alcuni autori sostengono che consonanza e fusione tonale siano fenomeni sovrapposti solo parzialmente. Huron (1991) ha analizzato gli scritti per pianoforte di Bach per investigare come tale autore utilizzasse gli intervalli armonici. L'analisi ha rivelato come Bach, per lo più, utilizzasse intervalli armonici consonanti piuttosto che dissonanti, con l'eccezione degli intervalli consonanti ben fusi come l'ottava e la quinta. In altre parole, secondo Huron (1991) la scrittura musicale di Bach fa intendere come tale autore prediligesse la consonanza ma, allo stesso tempo, tentasse per quanto possibile di evitare la fusione tonale per fare in modo che le varie voci polifoniche tipiche della sua musica fossero sempre ben distintamente udibili. In altre parole, secondo Huron, Stumpf sbagliava nel vedere una somiglianza tra consonanza e fusione tonale: suonare bene assieme a suonare fusi sono due cose differenti.

4.6 Gli intervalli armonici

In musica non si usano solo gli intervalli armonici, ovvero formati di note suonate contemporaneamente, ma anche gli intervalli melodici ovvero quegli intervalli ottenuti suonando una nota dopo l'altra in successione. La prima caratteristica notevole degli intervalli melodici rispetto a quelli armonici è la seguente: in musica, per lo più, vengono utilizzati intervalli melodici piccoli invece che grandi (Vos e Troost, 1989). Nella ricerca di Vos e Troost gli autori hanno analizzato

l'ampiezza degli intervalli utilizzati in un campione ampio di musica occidentale. I risultati che hanno osservato gli autori sono riassumibili in due punti. Anzi tutto gli autori hanno osservato come gli intervalli piccoli (ad esempio, l'unisono, la seconda minore, la seconda maggiore) fossero i più utilizzati in assoluto. Per contro, gli intervalli ampi (ad esempio la quinta, o l'ottava) erano poco utilizzati. Questo contrasta con quanto abbiamo visto per gli intervalli armonici. Gli intervalli armonici più utilizzati sono solitamente quelli ampi mentre quelli piccoli sono poco o per nulla utilizzati (Huron, 1991). Il secondo risultato osservato da Vos e Troost (1989) è che quando si suona un intervallo piccolo lo si suona per lo più in direzione discendente (ovvero re-do e non do-re), mentre quando si suona un intervallo ampio lo si suona per lo più in senso ascendente (ovvero do-la e non la-do).

Gli intervalli musicali melodici possono essere codificati in due modi diversi. Da un lato è possibile ricordare l'ampiezza di un intervallo, ovvero se l'intervallo è ampio oppure stretto. Dall'altro possiamo codificare la grandezza dell'intervallo in termini assoluti, ovvero l'esatto valore che separa le due note che compongono l'intervallo. Nel primo caso, ci ricordiamo ciò che in inglese viene spesso indicato come il *contour*, ovvero il contorno. In psicologia della musica la parola contorno viene utilizzata per indicare il profilo di saliscendi melodico generato da una melodia. La codifica degli intervalli musicali melodici (ma anche di quelli armonici) è diversa per i musicisti e per i non musicisti. I musicisti spesso possiedono l'*orecchio relativo*, ovvero, la capacità, date due note, di capire l'intervallo che descrivono le due note. Ad esempio se suoniamo un do ed un sol il musicista esperto è in grado di dire che le due note sono in intervallo di quinta pur senza sapere né il nome della prima nota, né il nome della seconda nota (questa caratteristica è appannaggio solo dei possessori del cosiddetto *orecchio assoluto*). Possedere l'orecchio relativo modifica il modo in cui il musicista codifica gli intervalli musicali. Tale differenza è evidente in quelle ricerche che hanno investigato se gli intervalli musicali sono percepiti categoricamente o meno. La percezione categorica è un fenomeno mutuato dalla percezione del parlato ma che può essere applicato a diversi domini cognitivi. Supponiamo di costruire dei suoni vocalici intermedi tra le sillabe /ba/ e /pa/ e di presentare tali suoni intermedi a degli ascoltatori e chiedere loro se odono /ba/, /pa/ o qualcosa di intermedio. Gli ascoltatori percepiscono o /ba/ o /pa/ anche nei casi intermedi, ovvero categorizzano i suoni ambigui riconducendoli all'una o all'altra delle due categorie. In altre parole, non riescono a percepire una consonante intermedia tra la /b/ e la /p/. Un fenomeno analogo si verifica per i musicisti. Burns e Ward (1978) hanno investigato il continuum tra la terza minore (do-re#) e la quarta (do-fa) creando una serie di intervalli che spaziavano da più piccoli della terza minore a più grandi della quarta. Tali intervalli venivano suonati a dei musicisti esperti a cui era richiesto di riconoscere l'intervallo. I musicisti dimostravano di percepire categoricamente i casi ambigui (ad esempio, gli intervalli compresi tra la terza minore e la terza maggiore) assegnandoli sistematicamente ad una (oppure all'altra) delle categorie disponibili. Invece lo stesso fenomeno non si manifesta nei non musicisti che riescono a percepire gli intervalli intermedi alternativamente sia come l'una che come l'altra delle categorie a disposizione. In altre parole, il non musicista è in grado di ascoltare un intervallo musicale inesistente (ad esempio uno intermedio tra una terza minore ed una maggiore) e percepirlo come differente sia dalla terza minore che dalla terza maggiore. Per contro, il musicista ascoltando lo stesso intervallo lo riconduce immediatamente

o ad una terza minore o ad una terza maggiore.

VEDI SLIDES

Il fatto che i musicisti percepiscano gli intervalli musicali in modo categorico non impedisce loro di essere (meglio dei musicisti) in grado di percepire l'intonazione fine degli intervalli musicali. L'intonazione fine degli intervalli musicali, infatti, può essere fatta in modo differente. L'intervallo di terza maggiore, ad esempio, può essere intonato in almeno tre modi differenti: quello naturale, quello temperato e quello pitagorico. Queste tre intonazioni differiscono di poco una dall'altra. Se intoniamo la terza maggiore in modo naturale le note dell'intervallo saranno separate da 386 centesimi di semitono. Se intoniamo in modo temperato le due note saranno separate da 400 cent. Infine, in modo pitagorico, le due note saranno separate da 408 cent. Le tre micro intonazioni differiscono di poco l'una dall'altra e, solitamente, un non musicista non è in grado di apprezzare le differenze tra queste tre microintonazioni. Per contro, queste microintonazioni vengono percepite in modo differente dai musicisti esperti. Tutte e tre queste microintonazioni sono corrette da un punto di vista musicale, tuttavia, ciascun musicista tende a preferire una particolare delle tre microintonazioni. La differenza nella percezione della microintonazione degli intervalli musicali tra musicisti e non musicisti è evidente sia quando il partecipante deve semplicemente percepire l'intervallo e riconoscerne l'intonazione sia quando al partecipante viene chiesto di produrre lo specifico intervallo, ad esempio con il metodo dell'aggiustamento. In una serie di esperimenti, Loosen (1994) ha chiesto ad un gruppo di pianisti, un gruppo di violinisti ed un gruppo di non musicisti di intonare in modo fine alcuni intervalli musicali. Il partecipante ascoltava una coppia di note e poteva modificare la frequenza della seconda nota della coppia. Il compito era quello di intonare il meglio possibile l'intervallo di terza maggiore (do-mi) di sesta maggiore (do-la) oppure di settima maggiore (do-si). I risultati hanno messo in luce come le intonazioni dei musicisti fossero molto più accurate di quelle dei non musicisti. Inoltre, i musicisti hanno dimostrato di preferire una specifica intonazione in funzione dello strumento suonato. I pianisti intonavano gli intervalli in modo temperato mentre i violinisti intonavano gli intervalli in modo pitagorico. Questa differenza probabilmente deriva dal fatto che i pianoforti sono solitamente intonati in modo temperato. Per contro, i violinisti possono intonare gli intervalli come meglio preferiscono e, solitamente, si considera l'intonazione pitagorica come la migliore tra le intonazioni per le esecuzioni melodiche, ed il violino è uno strumento prettamente melodico.

VEDI SLIDES

Le differenze nella percezione categorica e nella capacità di riconoscere o produrre le microintonazione degli intervalli musicali tra musicisti e non musicisti è il riflesso del lungo training a cui si sottopongono i musicisti. Tale training permetterebbe di memorizzare a lungo termine l'ampiezza specifica degli intervalli, così bene da poterli discriminare finemente anche nelle microintonazioni così come di categorizzarli (ovvero interpretarli) necessariamente entro specifiche categorie. Per quanto riguarda invece la memoria a breve termine degli intervalli musicali, le performance di musicisti e non musicisti sono meno differenti. Gli studi di Dowling (1978) e Dowling e Fujitani (1970) hanno investigato che

cosa memorizziamo a breve termine delle melodie che ascoltiamo. Nello studio di Dowling e Fijitani (1970) il soggetto ascoltava una breve melodia, che cominciava sempre dalla nota do, seguita dopo due secondi da una seconda melodia. L'esperimento era tra gruppi. Nel primo gruppo la seconda melodia poteva essere identica alla prima oppure essere casuale. Nel secondo gruppo, la melodia poteva essere identica oppure avere il contorno identico. Nel terzo, la seconda melodia poteva avere il contorno identico oppure essere casuale. I rimanenti tre gruppi affrontavano le medesime condizioni ma in tutti i casi la seconda melodia era trasposta, ovvero la melodia iniziava sempre da una nota differente da do. I risultati hanno evidenziato una differenza tra la performance nelle condizioni trasposte e nelle condizioni non trasposte. Quando la seconda melodia non è trasposta le due melodie vengono comparate