2 - Il suono come veicolo d'informazione

L'ambiente naturale è pieno di suoni, generati nell'ambiente stesso a causa di fenomeni fisici e, successivamente, biologici. In questo il *mondo sonoro* è assai diverso dal *mondo luminoso*, in cui la luce, nella stragrande maggioranza dei casi, proviene da una sorgente extraterrestre, il sole, e l'ambiente ne determina solo la propagazione in tutti i suoi aspetti: assorbimento, riflessione, diffusione, rifrazione. Se includiamo anche la radiazione infrarossa possiamo aggiungere le sorgenti ambientali di calore (vulcani, animali a sangue caldo), ma la precisazione non è particolarmente interessante, in quanto gli aggettivi "sonoro" e "luminoso" sono riferiti all'esistenza, negli organismi viventi, di sensori che rispondono alla luce (visibile) o al suono (udibile).

È probabilmente la diversa natura del meccanismo di trasduzione, fotochimico nel caso della luce, meccanico nel caso del suono, a determinare l'enorme differenza nell'intervallo di frequenze cui sono sensibili gli apparati visivi e quelli uditivi (nell'uomo circa un'ottava per la luce, dieci ottave per il suono); ma anche la diversa distribuzione in frequenza delle sorgenti luminose e sonore ha certamente influito nel determinare lo sviluppo di apparati sensoriali adatti a *cercare l'informazione dove essa è maggiormente presente*. La radiazione solare primaria (non filtrata dall'atmosfera) ha una distribuzione simile a quella di un corpo nero a 5500 gradi Kelvin: $f(\nu) \propto \nu^2 e^{-h\nu/kT}$ e quindi una elevata percentuale della radiazione è compresa nell'intervallo di lunghezze d'onda tra 400 e 800 nanometri. A questo occorre aggiungere che il filtro atmosferico, quello marino o lacustre e quello biologico (la parte trasparente dell'occhio dei vertebrati è costituita per la maggior parte di acqua) riducono ulteriormente la possibilità di utilizzare radiazione elettromagnetica al di fuori di quell'intervallo di frequenza.

Nel caso dei suoni invece, essi sono generati dall'ambiente stesso, e se pensiamo in termini fisici alle variazioni di pressione, le frequenze possibili vanno dai milionesimi di Hertz (i fenomeni meteorologici) ai megahertz o più dei fenomeni impulsivi, quali ad esempio il tuono; si tratta di una quarantina di ottave, e il fatto che un solo apparato, l'orecchio umano, riesca a coprirne un quarto, mi sembra un discreto successo.

Vediamo ora più in dettaglio quale informazione si può ricavare da una grandezza fisica (campo di pressione o campo elettromagnetico) che obbedisce all'equazione delle onde (di d'Alembert); principalmente la potenza che colpisce l'apparato di rivelazione, la direzione e la distanza di provenienza e la distribuzione spettrale, il tutto in funzione del tempo. Nel caso della luce la distribuzione spettrale primaria non dipende dal tempo, e l'intensità e la posizione della sorgente dipendono dal tempo in modo lento e regolare; inoltre la propagazione è praticamente in linea retta, dato che la lunghezza d'onda è di molti ordini di grandezza più piccola delle dimensioni degli oggetti che riflettono la luce. Pertanto le informazioni raccolte dagli apparati visivi possono dirci molto sulla posizione, la natura e il movimento di questi oggetti. Nel caso del suono invece la mancanza di una sorgente primaria fa sì che la variabilità nel tempo dipenda non solo dal movimento delle sorgenti sonore, ma anche dalle loro variazioni intrinseche durante l'emissione. Un analogo visivo di questa situazione potrebbe essere una scena in un bosco, in ogni punto del quale non solo vi sono oggetti in movimento, ma essi vengono illuminati in modo rapidamente variabile perché le foglie degli alberi agitate dal vento lasciano passare la luce solare in modo stocastico. Inoltre la lunghezza d'onda dei suoni udibili varia tra un centimetro e una diecina di metri, quindi gli effetti di diffrazione possono essere importanti rendendo la propagazione tutt'altro che rettilinea: il suono può facilmente girare intorno ad un ostacolo, perciò non è facile determinarne la direzione di provenienza.

Se confrontiamo un apparato visivo ed uno uditivo ci accorgiamo che entrambi fanno del loro meglio per raccogliere informazioni utili al benessere e alla sopravvivenza della specie o dell'individuo, ma adottano strategie differenti, forse proprio per le ragioni elencate in precedenza.

L'occhio ha un grande potere risolutivo spaziale (meglio di un primo), disponendo di un grandissimo numero di recettori retinici (circa 100 milioni), ma un cattivo potere risolutivo in frequenza (tre soli tipi di recettori cromatici), una elevata sensibilità alle variazioni temporali degli stimoli (nella regione periferica della retina), ma una scarsa capacità di analizzare singolarmente i movimenti degli oggetti, che vengono inseguiti e *foveati* per mezzo di opportuni movimenti oculari. Le variazioni di luminosità o di colore senza movimento vengono percepite solo su scale di tempo abbastanza lente e scarsamente memorizzate: è molto difficile leggere una successione di simboli che si sostituiscono rapidamente l'uno all'altro nella stessa posizione. Il potere risolutivo spaziale e la presenza di due occhi frontali dai movimenti controllabili consente anche una valutazione accurata delle distanze (stereovisione binoculare).

L'orecchio ha un mediocre potere risolutivo spaziale (soprattutto se i padiglioni esterni non sono mobili), un ottimo potere risolutivo in frequenza (i recettori interni sono circa 3000), una elevata capacità di percepire ed analizzare le variazioni temporali della pressione anche su scale di tempo brevi, e di memorizzarle. L'ascolto binaurale (stereofonia) permette anche di distinguere abbastanza agevolmente i segnali che provengono da sorgenti diverse senza confonderli tra loro.

Un'ultima osservazione per quanto riguarda la frequenza degli stimoli luminosi e di quelli acustici. La frequenza della luce visibile è compresa tra 0.375 e 0.75 milioni di miliardi di Hertz, e nessuno oggetto macroscopico (cellula, neurone) può vibrare con quella frequenza: la conversione fotochimica è un fenomeno quantistico, in cui un singolo fotone viene assorbito da una molecola del pigmento fotosensibile. La variabilità dell'intensità luminosa è limitata inferiormente dai tempi di latenza dei recettori, circa 1/20 di secondo. Le frequenze dei suoni udibili invece sono quelle compatibili con le varie parti dell'orecchio meccanicamente accoppiate con la membrana del timpano, ed il loro limite inferiore coincide, guarda caso, con l'inverso del tempo di latenza.

Succede allora che l'informazione uditiva (di un singolo orecchio) possa essere collocata in uno spazio a due dimensioni, il tempo e la frequenza, due variabili coniugate che possono essere collegate tra loro dalla trasformata di Fourier. A volere esser pignoli dovremmo dire che entrambe le variabili sono discretizzate, il tempo a intervalli di 1/20 sec, la frequenza a intervalli ottenuti dividendo il range di sensibilità (10 ottave) in circa 3000 intervalli (non necessariamente uguali).

Tuttavia la pressione sulla membrana del timpano, in quanto grandezza fisica oggettiva, dipende solo dal tempo in modo continuo, ed è possibile calcolarne la trasformata di Fourier, la quale dipende, in modo altrettanto continuo, soltanto dalla frequenza:

$$\pi(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} p(t) e^{i\omega t} dt. \tag{2.1}$$

Viene allora il sospetto che l'apparato uditivo non esegua proprio la trasformata di Fourier del segnale esterno, ma qualcosa che le somiglia: vediamo che cosa.

Gli estremi d'integrazione della formula (2.1) sono $-\infty, +\infty$, ed è ovvio che l'apparato uditivo non può aspettare la fine dei tempi, o più modestamente la fine del messaggio, per *Fourier-tradurlo*, visto che deve operare *in tempo reale*. Una proposta più ragionevole sembra essere quella di mettere come estremo superiore d'integrazione il tempo attuale, ottenendo così una funzione della frequenza e del tempo: questa presenta però l'inconveniente che gli ultimi eventi pesano molto poco rispetto a tutto il resto, diminuendo la capacità di reazione tanto più quanto più il messaggio è lungo. La soluzione migliore sembra quella della trasformata di Fourier *a finestra mobile*: si tratta di cancellare (con un certo criterio) gli eventi più remoti, in modo che la quantità d'informazione elaborata sia sempre la stessa (palinsesto). Il *certo criterio*, cioè la forma della finestra, può essere scelto in molti modi diversi, ma per quel che ci occorre basta immaginare una finestra quadrata, cioè gli estremi d'integrazione $t-\Delta t, t$, che ci fornisce una funzione del tempo, della frequenza, e della larghezza della finestra: nel nostro caso (umano) possiamo prenderla di 1/20sec.

È lo stesso tipo d'informazione che ci fornisce il sonogramma, un grafico in cui in ascisse c'è il tempo, in ordinate la frequenza, e l'annerimento è proporzionale al quadrato della trasformata di Fourier a finestra mobile: osservandolo ci rendiamo conto subito di quando e dove (in quali regioni di frequenza) è contenuta in prevalenza l'informazione di un certo evento sonoro.

La lettura di un sonogramma, visto che le intensità sono rappresentate da tonalità di grigio o per mezzo di un codice cromatico, non permette, a vista, di valutarle se non in modo qualitativo: più preciso è il sonogramma tridimensionale in cui le intensità sono rappresentate come altezze in un plastico e proiettate in modo prospettico sul piano.

Nella Fig.1 si vede uno pseudosonogramma, costruito numericamente, che rappresenta le prime quattro battute di "Fra Martino campanaro" eseguite prima da una voce grave e poi da una voce un'ottava sopra: entrambe le voci sono dotate di 8 armonici di uguale intensità. Un vero sonogramma non è così pulito e contiene molti più armonici, ma questo può dare un'idea. Proviamo a immaginare quale possa essere il *sonogramma ambientale* di una natura preindustriale, ottenuto con

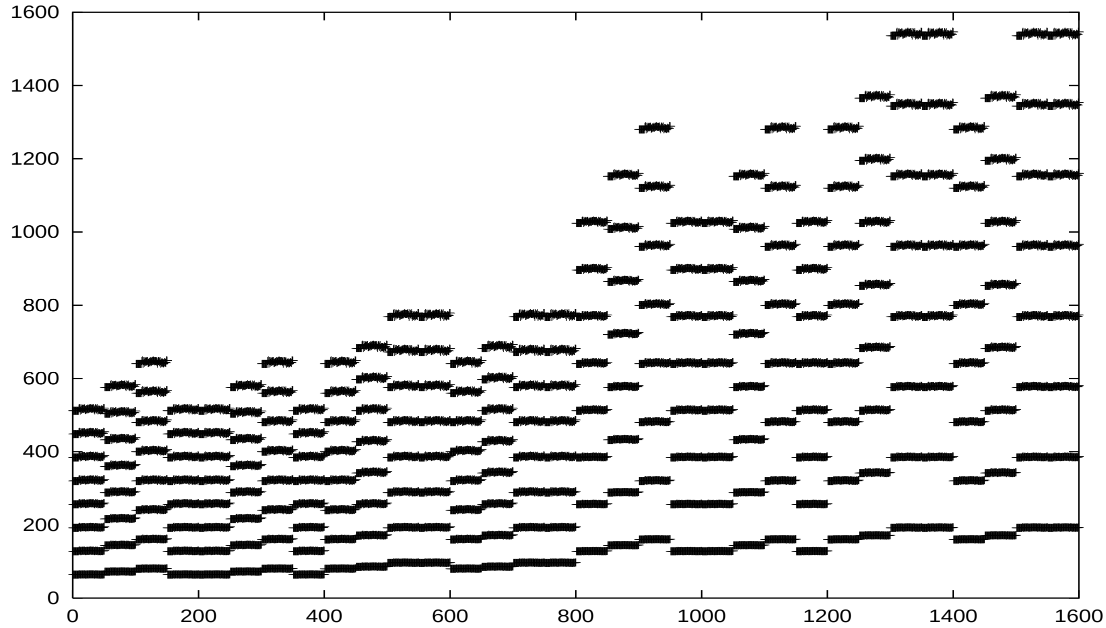

Fig. 1: Pseudosonogramma di "Fra Martino Campanaro"

una finestra mobile umana, cioè di 1/20 sec. Ci vengono alla mente suoni come quelli prodotti dal vento quanto incontra un ostacolo (fruscio dovuto alla vorticità, sibilo dovuto alle frequenze proprie di cavità, stormir di foglie), dall'acqua che scorre, cade come pioggia, grandine o cascata, sgocciola periodicamente da una foglia o si frange altrettanto periodicamente su una spiaggia o su una scogliera, dall'esplosione di un fulmine (il tuono e successivi rimbombi), da un terremoto o da un'esplosione vulcanica, da un fuoco, da una frana o valanga, dal battito lento o veloce delle ali di un uccello o di un insetto, dalla percussione sul suolo dei piedi di un animale, e infine dall'enorme varietà dei versi animaleschi. Ho volutamente elencato questi ultimi alla fine, per evidenziare il fatto che nella storia del nostro pianeta i suoni prodotti intenzionalmente costituiscono un fatto relativamente recente.

Se per un momento li escludiamo vediamo un sonogramma molto tranquillo, con una distribuzione spettrale pressoché continua e uniforme (rumore bianco) sia sulle basse che sulle alte frequenze, e una variabilità temporale scarsa, se non in occasione di eventi impulsivi distribuiti in modo stocastico, o di eventi periodici come quelli elencati prima. Se pensiamo al piano tempo-frequenza come ad una sorta di spazio delle fasi, possiamo dire che esso è occupato in modo abbastanza omogeneo, e che l'entropia è elevata. Ogni eventuale occupazione fortemente strutturata, sull'asse temporale o su quello delle frequenze, potrebbe attirare l'attenzione di un potenziale osservatore, allo stesso modo che la scoperta delle pulsar che emettono un segnale dotato di una sorprendente periodicità fece pensare a qualcuno che si trattasse di un segnale intenzionale inviato da qualche *omino verde*.

Facendo rientrare i versi animali nel nostro sonogramma vediamo che esso si arricchisce di segnali strutturati sia in frequenza che nel tempo: la loro distribuzione spettrale è prevalentemente discretizzata (spettro di righe) e presenta una forte correlazione temporale. L'occupazione dello spazio delle fasi è molto diversa e, soprattutto, localizzata, situazione alla quale corrisponde una diminuzione di entropia ed un incremento d'informazione.

Ci possiamo chiedere se sia nato prima l'uovo o la gallina, cioè l'orecchio o la voce. Probabilmente prima l'orecchio, visto che i suoni sono sempre esistiti, anche se non c'era nessuno in grado di ascoltarli, ma il primo che ne è stato capace ha ricavato sicuramente un vantaggio da questa nuova dote. Come potrebbe essere fatto un'orecchio primitivo e quali suoni potrebbe percepire? Se intendiamo "primitivo" dal punto di vista logico, un orecchio siffatto non può che essere un oggetto binario, cioè un dispositivo a soglia in grado di rispondere alle variazioni di pressione che la superano, indipendentemente dalla loro frequenza. Un esempio moderno di orecchio primitivo è l'antifurto delle automobili: questo esempio ci dà l'occasione di citare la latenza, quel periodo di tempo successivo alla risposta durante il quale il dispositivo non è in grado di rispondere ad un nuovo stimolo; se la soglia e la latenza sono ben calibrate (cosa che raramente accade per gli antifurto) il sistema è in grado di percepire anche le correlazioni temporali, purché sia dotato di *memoria*.

Che cos'è una memoria? È un dispositivo che colloca un'informazione che si sviluppa nel tempo (nel nostro caso) su un supporto che ne permetta una lettura, istantanea o temporalmente ordinata, in un tempo successivo alla fine dell'evento, per una durata più o meno lunga. Un esempio moderno è quello della registrazione di un evento temporale su un supporto di cera (incisione su rullo o su disco), di carta (pennino del sismografo o penna di chi scrive sotto dettatura), o elettromagnetico (polarizzazione del nastro magnetico): si tratta in ogni caso di una trasformazione del tempo in spazio (e della trasformazione inversa nel caso di rilettura del segnale memorizzato). In tutti questi casi il supporto spaziale e la punta scrivente sono in movimento l'uno rispetto all'altro. Com'è possibile realizzare una memoria uditiva in un cervello, visto che lì non c'è alcun moto? Con un insieme di linee di ritardo: supponiamo che dall'unico sensore dell'orecchio primitivo (binario) partano n fibre nervose di lunghezza $i \cdot dx; i = 1, 2, ..n$ che trasportano il segnale ad altrettanti sensori; se la velocità di trasmissione è v il tempo necessario per arrivare a destinazione sarà $t_i = i \cdot dx/v$; se l'evento sonoro ha una durata t_n esso si troverà *scritto* sugli n sensori; se la durata è maggiore, sui sensori risulterà solo l'ultima parte (palinsesto), che può comunque essere utile per evidenziare la presenza di eventuali correlazioni temporali all'interno dell'evento.

Possiamo allora affermare che un'orecchio accoppiato ad una memoria è in grado di coprire una regione rettangolare dello spazio delle fasi, con una base pari all'estensione temporale della memoria, ed un'altezza pari al range di frequenze cui è sensibile; la parte di sonogramma che cade in questa regione può esser analizzata con una precisione temporale dell'ordine del tempo di latenza, una precisione in frequenza pari al numero dei sensori indipendenti, ed una precisione in intensità legato al *numero delle gradazioni di grigio* che il sistema riesce a distinguere.

3 - L'uso del suono come mezzo di comunicazione

Proviamo a metterci nei panni di un essere dotato di un orecchio, più o meno primitivo, e di una memoria, più o meno lunga, e domandiamoci quali possono essere state le sue reazioni di fronte ai suoni ambientali: visto che la nostra precedente analisi del sonogramma puramente ambientale ce lo presentava come tranquillo, è legittimo pensare che l'essere in questione durante le fasi tranquille non sentisse niente, neanche le fluttuazioni stocastiche del rumore di fondo, e che si accorgesse solo degli eventi che, in un modo o nell'altro, emergevano dal fondo distinguendosi dalle fluttuazioni, o perché molto maggiori o perché fortemente correlati nella frequenza o nel tempo. Non ci interessa in questa sede se la sua reazione sia stata di paura, di curiosità o altro, quanto piuttosto quando (nello sviluppo filogenetico) si sia manifestata la consapevolezza di poter produrre intenzionalmente suoni percepibili. Ad un certo momento questo è successo, e quegli esseri hanno cominciato a scambiarsi non solo segnali ormonali (olfattivi) o ottici (visivi), ma anche acustici.

Mettiamoci ancora una volta nei loro panni ed esaminiamo quali caratteristiche deve aver un segnale acustico per essere effettivamente utile all'individuo ed alla specie. Deve emergere dal rumore di fondo, deve arrivare lontano, deve costare poco (in termini energetici), deve essere facilmente analizzabile da chi lo riceve, deve permettere di identificare il mittente, deve trasportare in breve tempo una elevata quantità d'informazione, deve essere robusto, cioè la perdita d'informazione durante il viaggio dev'essere minima; e magari deve essere protetto, cioè deve essere comprensibile solo da un gruppo di potenziali riceventi, gli *amici*.

Se l'orecchio del ricevente è del tipo binario e se non c'è memoria la sola possibilità è quella del "se ci sei batti un colpo", sperando che il colpo non coincida con un evento impulsivo ambientale che lo potrebbe mascherare e che sia abbastanza forte per poter arrivare lontano: supponendo che il livello del rumore (bianco) sia r, le sue fluttuazioni f e che l'intensità del segnale decada con il quadrato della distanza d, un segnale impulsivo (colpo) di intensità I verrà percepito fino ad una distanza d se $I > (r + 2f)d^2$; con questo tipo di meccanismo la quantità d'informazione che si può trasmettere e ricevere è scarsa, pari a 2: il colpo o c'è o non c'è. Se il sistema ricevente è dotato di memoria diventa possibile dare una struttura temporale al messaggio, in modo che la sua correlazione temporale, molto diversa da quella puramente stocastica, lo renda piú facilmente riconoscibile, e contemporaneamente si può trasmettere più informazione, precisamente 2^n, dove n è la lunghezza della memoria. Anche in questo caso si può fare un esempio moderno, l'alfabeto Morse, dove la struttura temporale è data dalle due diverse durate degli impulsi (punto e linea), ma anche dalla distanza tra l'uno e l'altro. Quest'ultima è largamente arbitraria, e caratterizza lo "stile di battuta" del telegrafista, che permette a chi riceve di riconoscerlo (identificazione del mittente), o addirittura di accorgersi se il mittente (riconosciuto) ha alterato leggermente il suo stile in modo da segnalare una situazione insolita, per esempio il fatto di trovarsi in uno stato di costrizione.

Val la pena di osservare che un linguaggio che occupa (come nel caso proposto) una piccola parte dello spazio delle fasi permette un metalinguaggio che si può manifestare attraverso una distorsione della porzione dello spazio delle fasi occupato. Vedremo in seguito molti altri esempi del genere.

Un incremento significativo della quantità d'informazione che si può trasmettere nella unità di tempo si può ottenere con la modulazione di ampiezza: i colpi non sono tutti della stessa intensità (zero o uno) ma possono avere m intensità diverse $(1, 2, ...m)$ purché l'orecchio del ricevente sia capace di distinguerle; la portata di questo sistema è quella del piú debole dei colpi, e la quantità d'informazione diventa m^n; l'inconveniente è che la robustezza diminuisce al crescere di m, soprattutto per i colpi forti, a meno che le intensità non crescano in modo esponenziale (per esempio ogni livello del 30% superiore a quello precedente).

Come si è visto una buona strategia è quella di occupare una parte molto piccola dello spazio delle fasi: un modo particolarmente efficace è quello di restringersi dal punto di vista della frequenza, in modo che l'energia del segnale possa competere vantaggiosamente con la componente del rumore nella stessa regione di frequenza. L'anatomia di molti animali permette l'emissione di suoni periodici, che hanno una distribuzione spettrale di righe di frequenza multipla intera di una frequenza fondamentale, pari all'inverso del periodo. Se l'orecchio del ricevente è "accordato" su quella frequenza o su una di quelle multiple, la comunicazione risulta enormemente facilitata; il codice potrebbe essere lo stesso di quello dei colpi.

È difficile immaginare un'orecchio che risuona esclusivamente ad una frequenza: è più verosimile pensare ad un risuonatore che risponde ad una banda centrata intorno ad una frequenza fissa con una certa larghezza di banda o, meglio ancora, ad un insieme di risuonatori con frequenze distanziate di una quantità confrontabile con la larghezza della banda di ciascuno, in modo da poter analizzare simultaneamente il contenuto in frequenza di ogni segnale. È il metodo adottato, in maniera molto grossolana, dai sistemi visivi cromatosensibili, i quali dispongono in ogni sito retinico di tre recettori (coni) dotati di curve di risposta parzialmente sovrapposte che coprono l'ottava della luce visibile: è in questo modo che noi percepiamo i colori, anche se sarebbe meglio dire che etichettiamo con aggettivi cromatici molto limitati l'enorme varietà delle distribuzioni spettrali. È difficile programmare geneticamente pigmenti fotosensibili, diversi da quelli che la natura è riuscita a produrre nel corso dell'evoluzione, per incrementare il potere risolutivo del sistema nell'ambito delle frequenze. È molto più facile programmare tanti diversi risuonatori acustici: basta variarne di poco le dimensioni per ottenere una scala molto ricca; nell'orecchio umano ve ne sono circa 3000.

Cerchiamo di ottimizzare le caratteristiche fisiche dei risuonatori in funzione della massima efficienza dell'apparato di rilevazione. Dato un oscillatore di massa m, costante di richiamo elastica k, costante di smorzamento s, forzato da una forza

$F(t)$, l'equazione del moto si scrive

$$m\frac{d^2x}{dt^2} + s\frac{dx}{dt} + kx = F(t); \tag{3.1}$$

l'equazione omogenea ($F(t) = 0$) permette di determinare la frequenza propria complessa ω che, per smorzamento subcritico ha una parte reale e una immaginaria date da

$$Re(\omega) = \omega_0\sqrt{1 - s^2/4mk}; Im(\omega) = s/2m; \omega_0 = \sqrt{k/m}. \tag{3.2}$$

Quando è presente un termine forzante periodico $Qe^{i\omega t}$ la soluzione di regime ha la stessa frequenza ed un'ampiezza

$$A = \frac{Q}{m({\omega_0}^2 - \omega^2 + i\omega s/m)}. \tag{3.3}$$

Per valori piccoli dello smorzamento s si osserva un comportamento risonante vicino alla frequenza propria ω_0, con una larghezza a metà altezza proporzionale a s/m, un tempo di decadimento inversamente proporzionale a s/m, ed una ampiezza inversamente proporzionale a $s\omega_0$

La selettività e la sensibilità del rivelatore richiederebbero uno smorzamento piccolo, ma la prontezza (capacità di rivelare con chiarezza un secondo segnale dopo la cessazione del precedente) richiede invece uno smorzamento critico. Inoltre può essere comodo che le ampiezze risultino indipendenti dalla frequenza di risonanza, in modo da avere una taratura globale. Come aggiustare i vari parametri (m, k, s) in modo da soddisfare le varie esigenze? Osserviamo intanto che lo smorzamento s, supponendo il moto in un fluido non viscoso né turbolento, è proporzionale alla superficie, e questa a sua volta può essere legata alla massa da una legge arbitraria, ma non assurda: una superficie proporzionale ad una qualche potenza λ non negativa della massa, minore o uguale a uno, sembra ragionevole. D'altra parte se lo smorzamento fosse inversamente proporzionale alla frequenza di risonanza avremmo un'ampiezza costante; ma la massa, se la costante elastica è costante, è inversamente proporzionale al quadrato della frequenza di risonanza. Mettendo insieme tutte queste possibilità otteniamo per la larghezza $\Delta\omega = d(k/\omega^2)^{\lambda-1}$ dove d è una costante. Solo il valore $\lambda = 1$ dà una larghezza indipendente dalla frequenza di risonanza; invece il valore $1/2$ dà luogo ad una larghezza proporzionale alla frequenza che suggerisce una distribuzione dei risuonatori omogenea in una scala logaritmica di frequenze.

La struttura dell'orecchio interno e della corteccia uditiva soddisfa in modo sorprendente questa esigenza: sia la collocazione delle cellule ciliate sull'organo del Corti che quella dei neuroni della cortecca uditiva seguono in buona approssimazione una scala logaritmica di frequenze. Entrambe possono essere schematicamente rappresentate come una tastiera, in cui ogni ottava ha lo stesso numero di tasti, non dodici ma circa 300.

Sembra quasi che alcuni fenomeni fisici, associati ad alcune considerazioni di carattere pratico, congiurino a dare un'aspetto logaritmico allo spazio delle frequenze: infatti se due esseri viventi vogliono comunicare con un segnale acustico ed hanno a disposizione una certa banda di frequenze, devono avere la certezza che il loro segnale sia robusto rispetto al variare (in modo compatibile con il loro comportamento) dei parametri caratteristici del loro codice di comunicazione. Che si tratti di una modulazione di ampiezza su di una portante di frequenza fissa, o di una modulazione di frequenza all'interno di una certa banda, la robustezza dev'esserci rispetto alle variazioni di frequenza dell'emittente dovute ad un eventuale "abbassamento di voce" o, più seriamente, all'effetto Doppler che si verifica quando l'emittente ed il ricevente sono in moto l'uno rispetto all'altro: se c è la velocità del suono e v quella del moto relativo, una frequenza ω viene percepita dal ricevente come $\omega(v) = \omega(1 + v/c)$, purché sia $v \ll c$. Si vede allora che allo scopo di permettere una ricezione completa su una frequenza ω quando la velocità massima tra i due è V la larghezza di banda dev'essere almeno $\Delta\omega/\omega \geq V/c$; analogamente le modulazioni di frequenza devono essere modulazioni della frequenza relativa.

Una velocità relativa pari all'un per cento della velocit'a del suono, cioè 3 m/sec (la velocità di un uomo che corre), produce una variazione di circa un sesto di semitono, certamente non molto rilevante; certi uccelli però raggiungono velocità relative dieci volte più grandi, e la variazione è quasi di un tono. Se l'informazione fosse contenuta nelle frequenze e non nei rapporti di frequenza, l'effetto Doppler la potrebbe modificare o addirittura distruggere.

Come si vede entrambe le esigenze conducono ad una struttura logaritmica nello spazio delle frequenze. Quel che rimane inevitabile è la trasformazione della regione dello spazio delle fasi delimitata dalla lunghezza della memoria e dall'intervallo di frequenza percepibile: quel rettangolo, in seguito all'effetto Doppler, si contrae in una direzione e si dilata nell'altra, pur mantenendo costante la sua area, e perciò quell'informazione che va a cadere nelle regioni periferiche rischia di andare perduta.

L'effetto Doppler non produce soltanto l'incremento (decremento) relativo delle frequenze, ma anche la riduzione (aumento) delle durate. Orbene la musica (o almeno una "certa" musica, quella basata sugli intervalli e sulle durate) è Doppler-invariante, come tutti hanno certamente verificato ascoltando un disco che gira ad una velocità maggiore o minore di quella prescritta: le frequenze variano, come anche la velocità di esecuzione, ma i rapporti di frequenza e di tempo (intervalli e ritmi) rimangono invariati, e il contenuto musicale è perfettamente riconoscibile.

C'è tuttavia, anche nella musica Doppler-invariante, qualcosa che invariante non è, e che in linea di principio permette di distinguere un segnale acustico *in movimento* dal corrispondente segnale *in quiete*. Immaginiamo l'esperimento seguente: un pianista dapprima registra un brano musicale come è stato prescritto dall'autore, successivamente lo registra eseguendolo a velocità dimezzata e un'ottava sotto (o a velocità doppia un'ottava sopra), ed infine le due registrazioni vengono poste a confronto, avendo cura di raddoppiare (dimezzare) la velocità della seconda registrazione, in modo che le frequenze e le durate tornino uguali a quelle della prima.

Chiunque, credo, è capace di riconoscere l'esecuzione genuina: vi sono infatti alcuni parametri del suono prodotto da uno strumento reale che non sono invarianti Doppler; essi sono la durata del cosiddetto transiente d'attacco e il timbro, legato tra le altre cose alle caratteristiche della tavola armonica. Una loro alterazione trasforma il suono originario in un suono *strano* nel senso che sembra prodotto da un pianoforte impossibile. Le stesse considerazioni si possono fare nei confronti degli altri strumenti: nel caso degli strumenti a fiato per esempio l'alterazione timbrica è dovuta al fatto che la velocità relativa tra suonatore ed ascoltatore è equivalente ad una variazione della lunghezza efficace della colonna d'aria, ma non della sua larghezza, ed è il loro rapporto che influisce sul timbro.

Naturalmente velocità relative molto piccole, inferiori al 10% di c influiscono sul timbro in modo impercettibile. Occorre tuttavia fare una precisazione a proposito dell'asserita Doppler-invarianza della musica: essa è limitata alla musica basata sul sistema temperato, in cui tutti i semitoni sono uguali e pari $2^{1/12}$. La musica basata su altri sistemi di accordatura (naturale, pitagorico, tono medio, ecc.) in cui i semitoni non sono tutti uguali, è solo approssimativamente Doppler-invariante. Un semplice esempio mostra il grado di approssimazione: nell'ambiente armonico di Do gli intervalli Si$\flat$-Do, Do-Re , Re-Mi valgono rispettivamente 8/7, 9/8, 10/9, mentre in ambiente temperato valgono tutti e tre $2^{1/6}$. L'incremento Doppler di un fattore 8/7 che porta il Si$\flat$ sul Do trasforma il valore 9/8 in 8/7, il 10/9 in 9/8. Per accorgersene occorre una sensibilità di 64/63 e di 81/80 rispettivamente, molto vicina all' 1%. Dubito che un orecchio musicale, anche raffinato, possa, o abbia potuto in passato, percepire tali sottigliezze, a meno che i suoni in questione non siano abbastanza prolungati e di ampiezza rigorosamente costante, come quelli emessi dalle canne d'organo, e per di più simultanei.

In uno dei precedenti capoversi ho accennato di sfuggita al timbro, una delle caratteristiche importanti del suono, che insieme al transiente d'attacco e al tempo di decadimento, permette di classificare in grande le famiglie di strumenti e in piccolo il singolo strumento. Ogni suono è costituito in genere da una frequenza fondamentale e da un insieme di soprasuoni (overtones) caratterizzati da ben definiti rapporti tra le loro frequenze e ampiezze con quelle del fondamentale. Per i suoni periodici i rapporti di frequenza sono numeri interi ed in tal caso i soprasuoni si dicono anche armonici; i rapporti di ampiezza invece possono variare ampiamente e caratterizzano il timbro del suono: un suono povero di armonici alti ha un timbro scuro, mentre uno ricco di armonici alti ha un timbro chiaro, ma gli aggettivi timbrici sono molto più variegati che quelli legati ad una sola dimensione, visto che spesso risultano udibili (sopra la corrispondente soglia di percettibilità) anche una ventina di armonici, corrispondenti a circa quattro ottave.

La capacità di modificare il timbro mantenendo costante l'intensità globale e la frequenza del fondamentale è comune a molti animali ed aggiunge un'ulteriore possibilità per arricchire il contenuto d'informazione per unità di tempo e caratterizzare ancora di più l'occupazione dello spazio delle fasi. Affinché questa capacità possa servire effettivamente a scambiarsi informazioni occorre che anche l'apparato ricevente sia capace di distinguere timbri diversi. Una valutazione di questa abilità nella specie umana può essere fatta esaminando le abilità musicali e linguistiche medie. La modulazione timbrica sembra il massimo della raffinatezza raggiunta in natura. Fisicamente ve ne sarebbe ancora una, la modulazione di fase, che non sembra venga utilizzata nel mondo animale, o meglio non viene utilizzata in fase di produzione di suoni, ma solo in fase di ricezione: questo mi sembra dovuto al fatto che le orecchie sono in genere due, e i segnali risultano sfasati tra l'una e l'altra, ma la sorgente è solo una e quindi lo sfasamento può dipendere soltanto dalla differenza di cammino dei due segnali. Uno sfasamento in emissione può risultare da uno sfasamento tra i vari armonici del suono, ma questo è difficilmente disaccoppiabile dalla variazione delle ampiezze (modulazione timbrica).

4 - Lo spazio delle percezioni uditive

Sulla base delle considerazioni precedenti voglio ora prendere in esame lo spazio delle percezioni udiitive per vedere se è possibile attribuirgli una qualche struttura di tipo matematico, per esempio una topologia e una metrica, cioè un criterio per definire una relazione di adiacenza e di distanza tra due percezioni, ed eventualmente quale; per questo è necessario definire correttamente gli stimoli e capire in che modo questi vengano codificati dall'organo sensorio, che nel caso naturale è l'orecchio, mentre nel caso artificiale è un qualsivoglia sistema di registrazione dei suoni.

Lo stimolo acustico è costituito dalle variazioni temporali della pressione dell'aria sulla superficie dell'organo sensorio, cioè dal valor medio di una funzione delle coordinate spaziali e del tempo, calcolato su quella superficie: per quello che ci interessa possiamo assumere il valore della pressione nel centro della membrana del timpano o del microfono e considerare lo stimolo acustico come una funzione soltanto del tempo $p(t)$; la sua rappresentazione grafica, a parte le distorsioni e i tagli introdotti dall'apparato di registrazione, è familiare a tutti. D'altra parte ciò che noi *udiamo* è ben diverso dal grafico: un grafico sinusoidale non viene *udito* come un'intensità oscillante nel tempo, ma come un suono puro di intensità costante e di altezza (nel senso musicale del termine) correlata alla frequenza, mentre un grafico complesso (sovrapposizione di molte sinusoidi) viene sentito a volte come un suono dal timbro particolare, a volte come due o più suoni simultanei, a volte come un suono di intensità variabile, a volte come una vocale e a volte anche come un fruscio, un rumore o comunque qualcosa non classificabile in termini musicali o alfabetici.

Val la pena di osservare che l'identificazione e la rappresentazione grafica degli stimoli sonori in termini alfabetici o musicali precede storicamente la conoscenza della natura fisica degli stimoli acustici e la conoscenza dei principi di funzionamento dell'orecchio, del nervo acustico e della corteccia uditiva. Quel che oggi sappiamo sull'argomento deve aiutarci a capire la struttura dello spazio uditivo. L'orecchio è essenzialmente un apparato che effettua qualcosa di simile alla trasformata di Fourier, cioè *calcola* i coefficienti dello sviluppo della $p(t)$ in combinazione lineare di funzioni sinusoidali pure: dal punto di vista matematico la trasformata di

Fourier può essere fatta conoscendo tutta la funzione (cioè per valori della variabile indipendente, il tempo, da meno infinito a più infinito) nel qual caso i coefficienti costituiscono un'infinità continua, o solo una parte di essa, per esempio in un intervallo di durata T, ma allora i coefficienti, che sono un'infinità numerabile, riguardano sinusoidi la cui frequenza è multipla di una frequenza fondamentale pari all'inverso del periodo T. L'orecchio non fa nessuna di queste due cose: si può dire piuttosto che vengono calcolati i coefficienti di un numero finito di frequenze distribuite all'interno di una certa banda passante, che qui schematizziamo nella banda compresa tra 20 e 20000 Hz: il risultato è quindi un "campionamento" della trasformata di Fourier, aggiornato circa 20 volte al secondo. Questo risultato può essere ottenuto con un insieme di "risuonatori", ognuno centrato su una particolare frequenza e con una certa larghezza di banda: è importante tener presente che questa larghezza è decisamente maggiore della "distanza" in frequenza tra due risuonatori adiacenti, per cui un suono puro ecciterà, in varia misura, tutti i risuonatori vicini a quello "giusto"; il fatto che noi "sentiamo" un suono puro (ma che cosa vuol dire, dal punto di vista percettivo?) può essere dovuto a meccanismi inibitorii tra i risuonatori vicini che restringono la banda, a tal punto che sulla corteccia uditiva l'eccitazione neuronale è molto più localizzata di quanto non sia l'eccitazione delle cellule dell'organo del Corti.

Siamo finalmente arrivati alla corteccia uditiva, ed è su questa che vogliamo ora esaminare la possibilità di dare una struttura topologica e metrica allo spazio percettivo.

Per ottenere questo risultato è necessario che lo spazio in questione possa essere in un certo senso *manipolato*, cioè sottoposto a trasformazioni controllabili: per fare un esempio consideriamo le percezioni visive, che sono il risultato dell'eccitazione dei neuroni di una area particolare della corteccia sensoriale, detta appunto corteccia visiva: quest'area è collegata alla retina per mezzo del nervo ottico, e si può pensare che l'immagine retinica venga *mappata* sulla corteccia visiva, forse con distorsioni, tagli o ricuciture: quando i nostri occhi si muovono con continuità l'immagine retinica cambia in modo prevedibile per mezzo delle leggi dell'ottica, ma non è detto che l'immagine corticale cambi nello stesso modo; sappiamo infatti che nella fovea i recettori sono molto più densi che nella regione periferica, e quindi un oggetto osservato con il centro dell'occhio coinvolge un numero di recettori molto più elevato di quelli coinvolti nella visione periferica dello stesso oggetto, e quindi le due immagini corticali sono molto differenti: questa è una variazione della metrica; una variazione topologica, tra retina e corteccia, potrebbe essere dovuta al fatto che nella retina esiste il cosiddetto *punto cieco*, in realtà una piccola regione, approssimativamente circolare, priva di recettori, mentre nella corteccia questo fatto non si verifica, e quindi due punti retinici lontani perché da parti opposte del buco potrebbero risultare vicini sulla corteccia. La consapevolezza di questi fatti è in parte legata alla possibilità di muovere gli occhi in vario modo e quindi di trasformare lo spazio visivo in sé stesso, e soprattutto di mettere queste trasformazioni in corrispondenza con il movimento degli occhi, o più in generale con quelle che io chiamo *manopole* per suggerire il loro ruolo come strumenti di controllo delle percezioni.

Quali sono le manopole di cui disponiamo per controllare lo spazio uditivo? Alcuni aninali sono capaci di muovere le orecchie, variando l'orientamento dei padiglioni auricolari, altri, come l'uomo, possono soltanto muovere la testa, e certamente questo tipo di movimenti può organizzare lo spazio percettivo acustico dal punto di vista della direzione di provenienza dei suoni, sfruttando le differenze di fase e d'intensità tra i segnali raccolti dalle due orecchie e la loro dipendenza dai movimenti del capo o dei padiglioni. Se cerchiamo manopole che riguardino la struttura dei suoni troviamo ben poco: avvicinandoci o allontanandoci con una certa velocità dalla sorgente sonora possiamo spostare lo spettro verso l'alto o verso il basso; altrimenti possiamo utilizzare una sorgente interna, la nostra voce, per modificare il suono e stabilire perciò una corrispondenza tra lo spazio uditivo e quello della produzione dei suoni.

L'udito e la vista sono però molto diversi dal punto di vista della distribuzione dell'informazione: nella visione l'informazione è soprattutto di tipo spaziale, anche se in alcuni animali e nell'uomo essa è rappresentata da un vettore nello spazio astratto dei colori, piuttosto che da una semplice intensità: la privazione delle percezioni cromatiche costituisce solo una menomazione secondaria, anche se talvolta fastidiosa, che non impedisce di cogliere gli aspetti più importanti dello stimolo visivo (si pensi alla ricchezza espressiva della grafica, della fotografia o della cinematografia in bianco e nero). Nell'udito accade il contrario: la perdita dell'informazione direzionale può essere grave per l'animale che deve avvertire la presenza del pericolo o dell'ostacolo con tecniche acustiche (il pipistrello, per esempio), ma è senz'altro irrilevante per chi, come l'uomo, dà soprattutto importanza al contenuto spettrale del segnale acustico.

Da questo punto di vista, purtroppo, quasi non disponiamo di manopole, e perciò le considerazioni fatte precedentemente ci porterebbero a concludere che non è possibile dare una topologia o una metrica a questo spazio, eppure la nostra esperienza, in particolare quella musicale, ci dice esattamente il contrario: siamo capaci di stabilire un ordinamento delle frequenze (di due suoni sappiamo dire in genere qual'è il più acuto e quale il più grave), abbiamo un criterio (che si è modificato nel corso dei secoli) per stabilire un indice di gradimento tra due suoni puri prodotti simultaneamente (consonanza e dissonanza), non abbiamo dubbi sul fatto che la *"é"* stretta e la *"è"* larga sono più vicine tra loro di quanto non siano la *"a"* e la *"i"*; in particolare se i due suoni prodotti simultaneamente distano di un'ottava (una frequenza è esattamente il doppio dell'altra) la senzazione è piuttosto quella di un unico suono con un timbro modificato piuttosto che quella di due suoni molto consonanti, e questa sensazione è in larga misura indipendente dalla frequenza stessa, come se un'ottava, o un qualsiasi altro intervallo tra due o più suoni, avesse le proprietà di un corpo rigido suscettibile di traslazioni senza deformazioni nello spazio delle frequenze. Perché? Io credo che la risposta a questa domanda vada ricercata in un ente matematico che io chiamo *matrice di correlazione a tempi uguali*, il quale riassume le esperienze uditive di ognuno di noi, o almeno di quelle avvenute nel corso del periodo in cui sono più attivi i meccanismi che presiedono alla costruzione delle connessioni sinaptiche: questa matrice potrebbe rappresentare in qualche modo la struttura delle connessioni nella corteccia uditiva; lascio ai neurobiologi il problema della loro effettiva realizzazione e lavorerò su di un modello.

La prima cosa da modellizzare è proprio la corteccia uditiva, che suppongo costituita da n neuroni i quali si comportano come risuonatori a banda stretta, in modo che un suono puro ne ecciti solo alcuni; la loro distribuzione (densità per intervallo di frequenza) può essere considerata una funzione arbitraria, anche se vi sono delle indicazioni sperimentali che suggeriscono una densità costante in funzione del logaritmo della frequenza; c'è inoltre una soglia, dipendente dalla frequenza, per cui si ha la massima sensibilità tra 3000 e 4000 Hz; dato che ci vogliamo occupare solo delle proprietà spettrali dei suoni possiamo lavorare su una sola corteccia anziché su due, senza perdita di generalità.

Lo stimolo acustico che inizialmente era stato identificato con la $p(t)$, pressione in funzione del tempo, viene convertito in una funzione della frequenza e del tempo, $f(\nu, t)$, con la convenzione che il tempo non è più una variabile continua ma viene discretizzata a intervalli dell'ordine del ventesimo di secondo, e così pure la frequenza, che va identificata con la frequenza di risonanza del neurone corticale i-esimo: il passaggio dalla $p(t)$ alla $f(\nu(i), t)$ comprende le distorsioni introdotte dal canale uditivo e dalla catena meccanica di trasmissione costituita dalla membrana del timpano e dagli ossicini, e la convoluzione con la funzione di risposta del neurone i-esimo. La matrice di correlazione (a tempi uguali) $G_{ij} \quad i, j = 1, 2, ...n$ si ottiene sommando su tutti i tempi (discretizzati) il prodotto dei valori di eccitazione dei neuroni i e j: $G_{ij} = \sum_t f(\nu(i), t) f(\nu(j), t)$; si tratta di una matrice reale e simmetrica (hermitiana quindi) i cui elementi non diagonali ($i \neq j$) ci danno informazioni circa la probabilità di eccitazione simultanea dei due neuroni; gli elementi diagonali invece sono proporzionali al valore quadratico medio dell'eccitazione di un singolo neurone. La matrice G dipende da tutte le esperienze uditive individuali, e ci si può aspettare perciò che individui diversi abbiano matrici diverse, ma che individui cresciuti in ambienti acustici simili abbiano matrici simili: si può dire in un certo senso che la matrice G riassume, sia pure in modo non ordinato cronologicamente, la "storia acustica" di ciascuno, e che da essa dipende la reazione di ciascuno alle successive esperienze uditive. Uno zero della G ci informa che quella coppia di neuroni non è mai stata eccitata simultaneamente; il rapporto tra G_{ij} e la media $(G_{ii} + G_{jj})/2$ ci dà la probabilità della loro eccitazione simultanea. Nulla vieta di prendere in considerazione funzioni di correlazione triple o di ordine ancora più elevato che coinvolgono l'eccitazione simultanea di tre o più neuroni: ne faremo un breve cenno in seguito a proposito della struttura dello spazio vocalico di una lingua e delle regole dell'armonia tradizionale.

Vediamo ora in che modo l'ambiente acustico può influire sulle caratteristiche della matrice G; le sorgenti sonore sono estremamente varie, ma possiamo raggrupparle in due grandi categorie secondo le caratteristiche dello spettro sonoro: sorgenti a spettro continuo (il suono che producono viene in genere chiamato rumore) e sorgenti a spettro di righe, ulteriormente classificabili in suoni armonici, se la frequenza delle righe è multipla intera di una frequenza fondamentale, e suoni anarmonici se la relazione tra le frequenze è di tipo diverso. Quel che però conta ai fini della matrice G sono i suoni percepiti istante per istante, che possono benissimo essere prodotti da più sorgenti attive simultaneamente: in generale lo spettro percepito è una sovrapposizione di una parte continua (fondo rumoroso) sul quale emerge lo spettro di righe, con una parte armonica e una anarmonica; sono i pesi relativi di queste tre componenti che determinano, in ultima analisi, le caratteristiche della matrice G.

La parte continua dello spettro contribuisce agli elementi di matrice della G in un modo che dipende lentamente dagli indici, mentre la parte di righe dà contributi particolarmente forti a quelle coppie di indici che corrispondono a due righe dello stesso spettro; se le relazioni tra le frequenze di due righe spettrali fossero distribuite a caso la matrice G risultante non sarebbe molto diversa da quella prodotta solo da uno spettro continuo. Però così non è: molti suoni naturali, in particolare la voce umana e quella di moltissimi animali, ma anche il ronzio prodotto dalle ali degli insetti, sono fenomeni periodici e hanno pertanto uno spettro armonico; ne segue che alla parte lentamente variabile della matrice G prodotta dai rumori e da una mistura statistica di suoni anarmonici (aperiodici), si sovrappone la parte molto strutturata, come vedremo fra un momento, generata dai suoni periodici. Ci occuperemo solo di quest'ultima, visto che il rumore non è strutturato, tenendo presente tuttavia che il rapporto segnale rumore ha poi la sua importanza nel determinare l'influenza che la parte strutturata della G avrà nel suo uso da parte dell'individuo e della sua simulazione computeristica.

Prima di andare avanti tuttavia sarà opportuno discutere brevemente alcune proprietà formali delle matrici di correlazione: abbiamo detto che si tratta di matrici reali e simmetriche, ma non tutte le matrici di questo tipo possono considerarsi matrici di correlazione: ricordando la definizione della G si vede che

$$G_{ii} + G_{jj} - 2G_{ij} = \sum_t (f(i,t)^2 + f(j,t)^2 - 2f(i,t)f(j,t)) \geq 0, \qquad (4.1)$$

trattandosi di una somma di quadrati; ne segue che

$$\mid G_{ij} \mid \leq [G_{ii} + G_{jj}]/2 : \qquad (4.2)$$

se vale il segno di uguale la correlazione tra i e j è la massima possibile. La G inoltre è sicuramente singolare (il suo determinante è nullo) se la somma su t non contiene almeno n contributi linearmente indipendenti, dove n è il numero dei neuroni corticali e la dimensione della matrice. Se gli n stimoli indipendenti costituissero una base (ortonormale) di uno spazio di Hilbert a n dimensioni allora la matrice G risulterebbe proporzionale alla matrice identità, che può essere vista anche come la somma dei proiettori sui vettori della base. Anche nel caso reale di stimoli qualsiasi la matrice G è una somma di proiettori sui vettori stimolo, ma questi vettori non sono in genere né ortonormali né linearmente indipendenti, per cui la G differisce dall'identità sia perché gli elementi diagonali non sono tutti uguali, sia per la presenza di elementi non diagonali dovuti alla presenza di correlazioni. È interessante vedere quali sono le proprietà di una matrice di correlazione prodotta da stimoli del tutto stocastici e incorrelati: ogni componente degli stimoli $f(i)$ è una variabile stocastica a valori tra 0 e F_i, il cui valor medio è $F_i/2$ e il valore quadratico medio $F_i^2/3$; se gli stimoli sono molto numerosi, trascurando le fluttuazioni statistiche, si ha $G_{ii} = {F_i}^2/3, G_{jj} = {F_j}^2/3, G_{ij} = F_iF_j/4$: ricavando F_i ed F_j dalle prime due relazioni si ottiene $G(i,j) = (3/4)\sqrt{[G_{ii}G_{jj}]}$, cioè gli elementi non diagonali sono tutt'altro che piccoli; in una generica matrice di correlazione il confronto tra G_{ij} e il valore precedente ci dice se la correlazione tra i neuroni i e j sia significativamente maggiore o minore di quella che ci si potrebbe aspettare in una situazione puramente stocastica.

Se, nel caso stocastico, tutti gli F_i fossero uguali a 1, gli elementi diagonali varrebbero 1/3, gli altri 1/4, per cui la G potrebbe essere considerata come la somma di una matrice con elementi tutti uguali a 1/4 più 1/12 della matrice identità: tenendo presente che l'identità proietta qualsiasi vettore sui vettori della base, cioè lo analizza nel modo più completo possibile, e che invece una matrice con elementi tutti uguali (la quale tra l'altro è singolare) non analizza altro che la somma delle componenti di un vettore, poiché lo proietta "nello stesso modo" su ogni vettore di base, si può capire perché è importante che il livello di "rumore stocastico" sia il più basso possibile, per aumentare il potere analizzante della matrice di correlazione. Conviene notare anche che un insieme di stimoli a componenti non negative non può dar luogo ad una G proporzionale all'identità, se non nel caso in cui ogni stimolo sia dotato di un'unica componente diversa da zero, ma in tal caso lo spazio percettivo non può acquistare né una topologia né una metrica.

Un suono periodico di periodo T può essere rappresentato dal suo sviluppo in serie (discreta) di Fourier secondo la formula $f(t) = \sum_k A_k \sin(k\nu t + \phi)$, dove ν è la frequenza fondamentale pari a $2\pi/T, k = 1, 2...$ e A_k è l'ampiezza relativa alla k-esima armonica; il suo contributo alla matrice G è diverso da zero solo per quelle coppie di indici che corrispondono a frequenze entrambe presenti nello spettro e vale $g_{ij} = A_k\delta(\nu(i), k\nu) \cdot A_l\delta((\nu(j), l\nu)$, dove $\delta(a, b)$ è il simbolo di Kronecker o la delta di Dirac. La rappresentazione grafica di questo contributo è presentata nella Fig.2, dove gli indici di riga e di colonna della matrice sono proporzionali al logaritmo della frequenza; un suono dello stesso tipo, ma con un diverso periodo, dà un grafico simile ma traslato parallelamente alla diagonale principale di una quantità proporzionale al logaritmo del rapporto tra le frequenze: a questa rappresentazione manca l'informazione relativa alle ampiezze dei vari armonici, che richiederebbe un grafico tridimensionale. Se i due suoni vengono presentati in successione il contributo complessivo è dato dalla somma dei contributi, e il grafico dalla sovrapposizione dei grafici (Fig.3) , ma se la presentazione è simultanea a questo contributo si deve aggiungere quello della correlazione incrociata, rappresentato dalla Fig.4. Supponiamo ora che questi suoni, presentati in successione, siano talmente numerosi da coprire, con le loro frequenze fondamentali, un ampio intervallo, per esempio l'intera banda

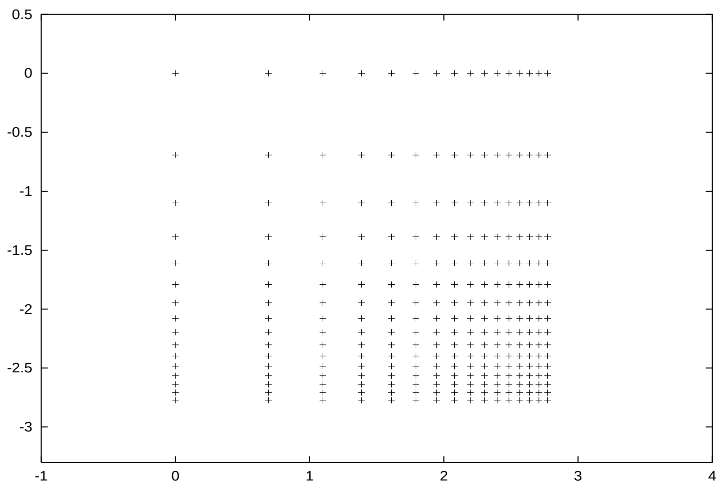

Fig. 2: Matrice di correlazione prodotta da un suono periodico con 12 armonici.

audio: il risultato è costituito dalla sovrapposizione di tutti i contributi (simili) dei singoli suoni, che viene mostrato qualitativamente dalla Fig.5.

La struttura del frammento di matrice di correlazione mostrato nella Fig.5 evidenzia l'invarianza traslazionale nel logaritmo della frequenza, e può essere accettata come la *vera matrice di correlazione neurologica* se ricordiamo che le cellule ciliate dell' Organo del Corti presentano un densità quasi costante nel logaritmo della frequenza. Se si fosse tenuto conto dell'ampiezza dei vari armonici, che comè noto decresce con l'ordine dell' armonico, la diagonale principale e le righe isolate sarebbero risultate molto più intense delle altre: una rappresentazione di una riga di questa matrice verrà presentata nel capitolo successivo.

Se accettiamo questo modello di organizzazione corticale ci possiamo domandare quale può essere effettivamente la matrice di correlazione generata da un ambiente acustico "ragionevole": ci sarà certamente un fondo continuo, dovuto al rumore stocastico o, ciò che è del tutto equivalente, alla presentazione simultanea di un grandissimo numero di suoni periodici, poi una parte di righe del tipo di quella presentata nelle figure, e finalmente una parte costituita da righe in posizione "anomala" prodotte da suoni aperiodici e da punti o frammenti di righe prodotti dalla presentazione simultanea di pochi suoni periodici e/o aperiodici.

In un ambiente acustico dominato da suoni periodici, quale potrebbe essere quello abitato una comunità umana non industrializzata, eventualmente fornita di animali, la matrice, a parte un fondo continuo, assomiglierà a quella della fig.5: con quale risultato? Che neuroni corrispondenti a frequenze multiple di un comune fondamentale (e quindi in rapporto razionale del tipo m/n) sono, per così dire, abituati ad essere eccitati simultaneamente, tanto più quanto più gli interi m, n sono piccoli, e perciò l'eccitazione di uno solo di essi comporta uno stimolo anche di quelli ad esso correlati: questo fatto, da solo, può spiegare la maggior parte delle regole

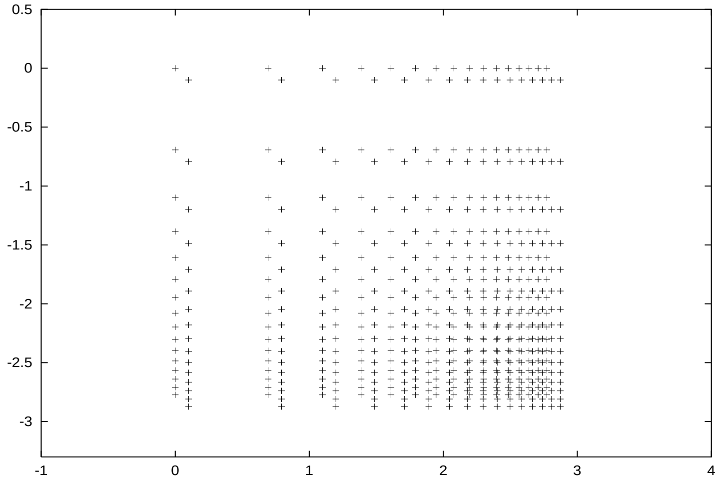

Fig. 3: Matrice di correlazione prodotta da due suoni periodici in successione.

dell'armonia tradizionale, e anche la circostanza che le scale musicali sono, appunto, scale e non piani inclinati, cioè che i suoni sono in numero piccolo (5,7,12,31 a seconda delle culture musicali) e formano tra di loro intervalli relativamente semplici. Questo non impedisce ovviamente la pratica del *glissando*, sia in termini vocali che strumentali, in cui si concretizza la consapevolezza che l'universo dei suoni si può appoggiare ad un continuo unidimensionale, continuo che però, attraverso l'esperienza uditiva, viene distorto e attorcigliato come ora cercherò di mostrare. Con quale criterio si stabilisce la *distanza* tra due suoni? Uno è quello legato alla topologia unidimensionale, sulla quale la distanza tra due suoni (puri) è data dalla differenza tra le rispettive frequenze, o meglio dal logaritmo del loro rapporto, per cui se i due suoni coincidono la distanza è in ogni caso zero, se distano di un'ottava la distanza è uguale alla frequenza del suono più basso (oppure $\log 2$), e così via. Ma c'è un altro criterio, di carattere musicale, per cui invece due suoni che distano di un'ottava sono molto più *vicini* tra loro rispetto a due che distano di un semitono, per non parlare della distanza tra due suoni il cui rapporto di frequenza sia dato da numeri primi molto grandi o sia addirittura un numero irrazionale. A voler essere pignoli si potrebbe aggiungere un terzo criterio, a carattere molto soggettivo, basato sulla *gradevolezza* o sensazione di consonanza che si prova durante il loro ascolto simultaneo: questo criterio è importante, dal punto di vista musicale, quando i suoni vengono prodotti da strumenti reali, con i loro bravi armonici, i cui pesi relativi sono diversi da uno strumento all'altro: lo stesso intervallo può risultare gradevole quando a produrlo è una coppia di strumenti, sgradevolissimo per un'altra coppia. L'uso accorto della matrice di correlazione permette di giustificare in larga misura sia il secondo che il terzo criterio.

Se si dà una rapida occhiata alla Fig.6, che rappresenta l'andamento della consonanza in funzione della distanza tra due suoni (valutata con il criterio numero tre) si

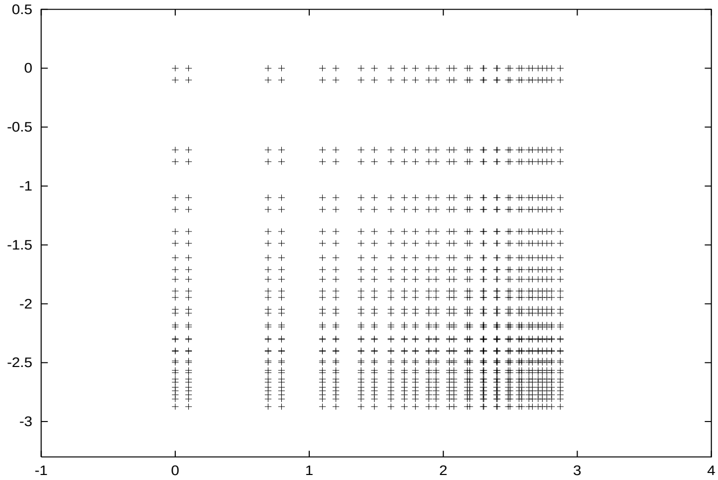

Fig. 4: Matrice di correlazione prodotta da due suoni periodici simultanei.

vede che il suono più vicino a un suono di riferimento è quello che con questo forma un intervallo di ottava, seguito dalla quinta, dalla quarta, dalla terza maggiore, e così via, mentre estremamente lontani (consonanza minima o addirittura nulla) sono intervalli piccoli come il semitono (e dintorni) o vicini all'ottava, come la settima maggiore.

È difficile inventare un'organizzazione matematica formale dello spazio percettivo che renda conto simultaneamente di almeno due dei criteri di valutazione della distanza: è ovvio che una dimensione non è sufficiente, visto che la distanza di ottava è grande secondo un criterio e piccolissima secondo l'altro; servono più dimensioni, non necessariamente continue, e una prima rappresentazione potrebbe essere quella, già spesso usata , dell'elica cilindrica (fig7), in cui l'asse del logaritmo della frequenza viene *arrotolato* sulla superficie di un cilindro, in modo che un giro corrisponda ad un'ottava: in questo modo la vicinanza delle ottave viene rispettata, ma non le relazioni di vicinanza o gradevolezza di altri intervalli.

Un altro metodo, anch'esso basato sull'elica cilindrica, che cerca di dare una rappresentazione spaziale della matrice di correlazione, consiste nella costruzione di una serie di linee che collegano sul mantello del cilindro coppie di punti sull'elica che distano di un intervallo privilegiato, cioè l'ottava, la dodicesima,ecc., in breve il fondamentale con i propri armonici e questi tra loro, almeno quelli di ordine più basso, come schematizzato in fig 7. Ad ogni punto del continuo dell'elica è associato un insieme numerabile (in sostanza le frequenze in rapporto razionale) che può essere immaginato come un reticolo cristallino, reso finito dai limiti dell'orecchio: in questo modo le distanze tra i suoni *appartenenti allo stesso reticolo* possono essere valutate in base al numero di legami ed alla loro forza, e questo è un modo di quantificare il criterio numero due; la distanza tra suoni appartenenti a reticoli diversi invece può essere misurata solo sulla topologia unidimensionale, cioè con il criterio numero uno.

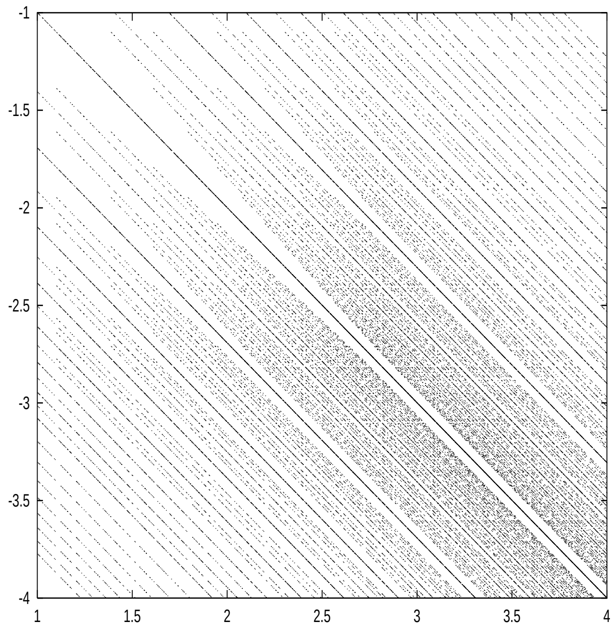

Fig. 5: Matrice di correlazione prodotta da 400 suoni periodici in successione.

Una semplificazione della rappresentazione elicoidale si può ottenere proiettando l'elica su un piano perpendicolare all'asse del cilindro, in modo da ottenere una circonferenza la cui coordinata angolare corrisponde alla frequenza, modulo un'ottava: per la precisione l'angolo, misurato in radianti, è proporzionale alla mantissa del logaritmo in base 2 del rapporto tra la frequenza e una frequenza di riferimento $\phi(\nu) = 2\pi(\log(\nu/\nu_o) - \int(\log(\nu/\nu_o))/\log 2)$, da cui si vede come suoni che differiscono per una o più ottave siano rappresentati dallo stesso angolo. Su questa circonferenza ci si può muovere con continuità (glissando) oppure a salti di semitono, di quinta o di qualsiasi altro intervallo.

Se si vuole evidenziare graficamente la duplice possibilità di moto continuo o discreto si può ricorrere ad una rappresentazione un po' più complicata che utilizza la superficie di un toro: su questa si può tracciare una linea continua chiusa che non incrocia sé stessa sulla quale sono rappresentate tutte le frequenze contenute in un ottava; i salti di quinta, che rappresentano la massima consonanza (dopo l'ottava) collegano punti di quella linea che hanno lo steso valore di una delle due coordinate angolari del toro.

Dovrebbe essere chiaro da quanto detto fin'ora che il concetto di distanza ha senso solo quando si tratta di suoni puri, rappresentati comunque da punti. Un suono dotato di armonici occupa con pesi diversi alcuni punti di un reticolo, per cui la sua distanza da un altro suono può essere considerata pari alla distanza tra i fondamentali, misurata con il criterio uno o due, oppure valutata in modo completamente diverso, considerando i suoni come vettori in uno spazio lineare:

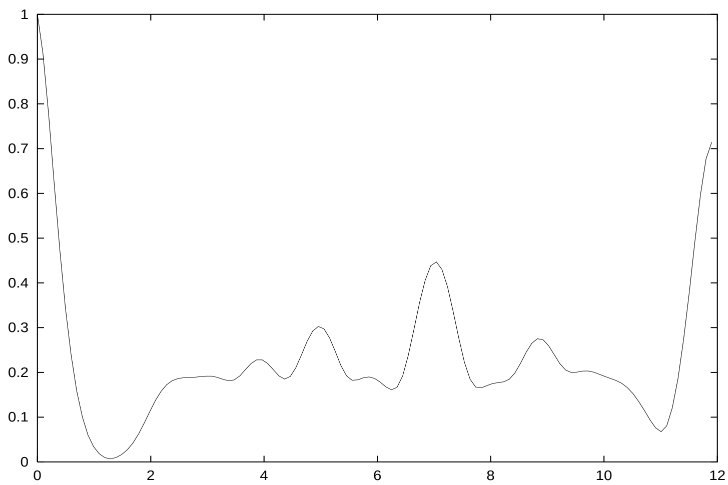

Fig. 6: Andamento della consonanza all'interno di un'ottava.

allora il loro prodotto scalare è proporzionale al coseno dell'angolo, e la tangente di quest'angolo può essere considerata una ragionevole misura della loro distanza: maggiore è il numero degli armonici che essi hanno in comune minore è la loro distanza. Questo criterio è oggettivo, anche se presenta alcuni inconvenienti: due suoni puri non coincidenti hanno sempre distanza infinita, dato che i loro armonici non vengono espressi; meglio allora usare il criterio soggettivo che si serve della matrice di correlazione: al prodotto scalare $< s_1 \mid s_2 >$ tra i due suoni si sostituisce l'elemento di matrice della matrice di correlazione $< s_1 \mid G \mid s_2 >$: anche due suoni puri diversi, se le rispettive frequenze sono in rapporto razionale, vengono *sentiti* come appartenenti allo stesso reticolo dalla matrice di correlazione, e quindi la loro distanza non è infinita.

Come si vede la rappresentazione delle frequenze su di una superficie, cilindrica o toroidale che sia, non corrisponde a una vera bidimensionalità dello spazio dei suoni, che rimane unidimensionale per i suoni puri, e infinitodimensionale per quelli dotati di armonici. Esiste tuttavia una caratteristica del nostro sistema percettivo che permette di considerare bidimensionale questo spazio, restituendo così un senso alla rappresentazione cilindrica: noi riusciamo in qualche modo a separare il tono (il nome della nota o l'angolo sulla circonferenza descritta poco fa) dall'altezza: per i suoni puri questo è possibile a salti di ottava, ma per i suoni composti può essere realizzato con continuità, anche se questa affermazione può sembrare paradossale. È possibile realizzare una successione di suoni (l'equivalente acustico delle scale di Escher) che percorra un scala (diatonica o cromatica) al termine della quale la nota con lo stesso nome della nota di partenza non si trovi un'ottava al di sopra o al di sotto, bensì esattamente alla stessa altezza: per ottenere questo risultato basta costruire con un sintetizzatore la successione dei suoni alterando ad ogni passo le ampiezze degli armonici (e questo altera ovviamente il timbro) in modo da distribuire su tanti piccoli passi la grande alterazione necessaria per compensare il salto di ottava. La rappresentazione grafica di questa scala è quella di una successione (discreta o continua) di punti che percorrono il mantello del cilindro ad un'altezza costante, cioè su di una circonferenza.

Una volta imparato il trucco si possono costruire tanti percorsi "impossibili" quali ad esempio il glissando di ottava, in cui il nome della nota rimane costante ma varia l'altezza (una retta verticale sul cilindro), o le scale ambigue, in cui i nomi delle note si susseguono in ordine crescente (decrescente), ma alla fine ci si trova un'ottava più in basso (alto) (un'elica sinistrorsa anziché destrorsa), e tutti gli infiniti percorsi che si possono costruire combinando questi elementi.

Con questa tecnica tutti i punti della superficie cilindrica corrispondono a possibili suoni, diversi tra loro, mostrando in modo esplicito che lo spazio percettivo è (almeno) bidimensionale.

Vorrei concludere questo capitolo con una osservazione sul problema, antico quanto la musica stessa, dell'incommensurabilità tra le quinte e le ottave, dovuto al fatto che sette ottave (sette giri sull'elica cilindrica) sono solo approssimativamente uguali a dodici quinte, con tutte le conseguenze che questo ha comportato nella storia dell'accordatura degli strumenti: il continuo unidimensionale, i reticoli dei razionali, sono tutte bellissime cose dal punto di vista matematico, ma affidarsi completamente alla matematica quando c'è di mezzo anche una realtà fisica può essere pericoloso. La frequenza di un suono di durata limitata, indipendentemente dal fatto che venga misurata da uno strumento tecnologico o riconosciuta da un orecchio (musicale), non è rappresentata solo da un numero reale (razionale o irrazionale), ma anche da un'altro numero che rappresenta l'errore, o l'incertezza della misura, il cui risultato deve quindi essere correttamente espresso da $\nu \pm \Delta\nu$: il rapporto tra le frequenze di due suoni è anch'esso affetto da un errore, una banda di incertezza intorno ad un valore medio, entro la quale cadono comunque infiniti numeri razionali; allora i reticoli generati dai due suoni avrebbero sempre qualche punto in comune, o più correttamente qualche regione (punto più errore) in comune; tuttavia questa regione potrebbe trovarsi molto lontano sull'elica, in una zona in cui l'orecchio è sordo, ed ecco allora che i reticoli di punti matematici ben distinti l'uno dall'altro, in parte confusi da effetti fisici, vengono nuovamente separati da cause fisiologiche.

A proposito dell'accordatura quindi, il problema secondo me ha un senso da un punto di vista teorico, mentre dal punto di vista pratico è rilevante soprattutto per gli strumenti a tastiera a suono prolungato e, aggiungo malignamente, dall'accordatura stabile: se uno strumento come il clavicembalo o il liuto va riaccordato più volte nel corso della stessa esecuzione, poco importa se la stonatura dipende dall'uso di un particolare sistema di accordatura o dal fatto che le chiavi non tengono bene. Per gli strumenti a fiato o ad arco, o per la voce, il problema quasi non esiste, perché è possibile, variando momentaneamente la tensione di una corda o spostando con continuità il dito, oppure variando l'intensità del soffio o la lunghezza del tubo, adeguare l'intonazione di qualsiasi nota alle esigenze del momento. In conclusione sette ottave *sono* uguali a dodici quinte, *entro gli errori sperimentali*, salvo casi particolari in cui *bisogna fare attenzione*.

Il sistema temperato, se lo si vuole rappresentare sul solito cilindro, comporta una leggera distorsione del reticolo associato ad ogni punto, per cui alla fine l'ottava, che è un cerchio, viene divisa in due parti uguali dalla quarta aumentata, in tre dalla terza maggiore, in quattro dalla terza minore, in sei dalla seconda maggiore (scala esatonale di Debussy) in dodici dalla seconda minore (scala cromatica, dodecafonia, Schoenberg): in questo modo viene resa esatta una simmetria che prima era soltanto approssimata; il prezzo pagato è quello della rinuncia ai rapporti *esattamente razionali* degli intervalli musicali.