

## 10 - La voce umana

Lo strumento a fiato senz'altro più importante è la voce, ma è anche il più difficile da trattare in modo esauriente in queste brevi note, a causa della sua complessità. Vediamo innanzitutto le caratteristiche schematiche del processo di emissione dei suoni da parte dell'apparato fonatorio umano: esso consiste in un tubo (il cosiddetto tratto vocale) lungo in media 17cm, aperto a una (o due) estremità, la bocca e il naso, che può essere eccitato da un meccanismo ad ancia (le corde vocali) o da un getto d'aria che attraversa l'apertura orale, opportunamente ristretta, in uscita o in entrata (il fischio); si potrebbe aggiungere un terzo meccanismo in cui il respiro può diventare rumoroso (ma non sonoro) nell'attraversamento della laringe; il tubo ha una sezione variabile che può essere controllata in vari modi, modificando l'apertura delle labbra, il volume del cavo orale e la posizione della lingua, e inoltre coinvolgendo o meno il tratto nasale. Nella cosiddetta posizione neutra il tubo si comporta in modo simile ad un tubo cilindrico di lunghezza equivalente eccitato da un'ancia: le sue frequenze proprie (massimi dell'impedenza d'ingresso) vengono convenzionalmente fissate ai multipli dispari di 500Hz. Oltre ad avere la sezione controllabile il tubo ha pareti molli e capaci di scambiare anche vapore acqueo con la colonna di gas: per questo motivo occorre tener conto di un'attenuazione dell'onda di pressione nel corso della sua propagazione attraverso il tratto vocale: come conseguenza le risonanze diventano più larghe. L'eccitazione prodotta dalle corde vocali consiste in una pressione che è una funzione periodica del tempo, con una frequenza che, dal basso profondo al soprano sopracuto, varia tra 50Hz e 2000Hz, ed ha uno spettro armonico più o meno ricco, secondo che il movimento di apertura e chiusura della laringe è brusco (dente di sega) o morbido (sinusoide); un singolo cantante ben addestrato può avere un range di frequenze di due o più ottave. La larghezza delle risonanze permette di utilizzare tutte le frequenze di cui sono capaci le corde vocali.

Il controllo della fonazione, o del canto, si basa su quattro meccanismi, le modulazioni di ampiezza, di frequenza, di spettro e di timbro: la prima si ottiene variando la pressione diaframmatica (e quindi il flusso dell'aria attraverso le corde vocali), la seconda variando la tensione delle corde vocali, la terza controllando la durezza del suono, la quarta modificando la forma del tratto vocale. mentre ognuna delle prime è riconducibile alla variazione di un unico parametro, l'ultima è un po' più complessa, soprattutto nella sua modellizzazione fisica: le molteplici variazioni della sezione del tratto vocale (labbra, bocca, lingua, naso) hanno come risultato quello di spostare i massimi di impedenza i quali (allargati dai fenomeni di assorbimento) funzionano come filtri passabanda del segnale primario prodotto dalle corde vocali: questi filtri prendono il nome di "formanti" e determinano il timbro del suono emesso, a patto che il segnale primario prodotto dalle corde vocali sia abbastanza ricco di armonici; in particolare sono le prime due formanti, e in misura molto minore la terza, che con i loro spostamenti permettono di ottenere una grande varietà di timbri, che vengono percepiti come le diverse vocali di una lingua. Lo spostamento delle formanti è notevole: la prima può variare tra 250Hz e 750Hz, la seconda tra 600Hz e

2000Hz, la terza tra 1700Hz e 3000Hz per le voci maschili, mentre nelle voci femminili esse sono leggermente più alte, dato che il tratto vocale è mediamente più corto. Un modello relativamente semplice del tratto vocale consiste in un certo numero (da quattro a sette, a seconda della flessibilità che si vuole ottenere) di settori di lunghezza assegnata e sezione costante (il cui valore può essere modificato) in ciascuno dei quali la pressione è rappresentata da un'onda progressiva e una regressiva: imponendo le solite condizioni all'estremo aperto verso l'esterno (bocca) e a quello collegato all'ancia, nonché le condizioni di continuità al confine tra un settore e l'altro, si trova l'andamento dell'impedenza d'ingresso e quindi la posizione e la larghezza delle formanti.

A parità di formanti la massima efficienza nella modulazione timbrica si ottiene quando il segnale primario è ricco di armonici, sia nel senso del loro numero che in quello della loro densità nella regione di frequenze in cui le formanti agiscono; è per questo motivo che una voce maschile bassa, con una frequenza intorno ai 100Hz è più intelligibile di una femminile acuta, intorno ai 500Hz: la prima infatti possiede circa 30 armonici nella regione delle prime tre formanti (fino a 3000Hz), mentre la seconda soltanto 6, i quali portano una minor quantità d'informazione. Il massimo della densità d'informazione, anche se di debole intensità, è quello portato dalla voce soffiata, visto che il suo spettro è continuo.

Voglio presentare ora alcuni risultati numerici che mostrano, sia pure in modo grossolano, da una parte l'allargamento delle formanti dovuto ai fenomeni di assorbimento, dall'altra il loro spostamento dovuto alla variazione delle sezioni. Nel primo caso considero un tubo cilindrico di raggio  $a$  costante e lunghezza  $L$ , eccitato da un'ancia, in cui l'assorbimento delle pareti fa sì che il numero d'onde  $k$  diventi complesso  $k = \omega/v - i\alpha$ , dove la velocità differisce leggermente dalla velocità del suono  $c$  in ambiente aperto e non assorbente  $v = c(1 - \cos t/(a\omega^{1/2}))$ , mentre  $\alpha = a\sqrt{\omega}L$ ; per un tubo idealmente aperto in cui  $Z_L = 0$  si ha

$$Z_{in} = Z_o \frac{\tanh \alpha L + i \tan(\omega L/v)}{1 + i \tanh \alpha L \tan(\omega L/v)}. \quad (10.1)$$

Nella Fig.1 è graficato l'andamento delle prime quattro formanti (il logaritmo del modulo quadro dell'impedenza d'ingresso) supponendo  $v = c$  e variando  $\alpha$ : è chiaramente visibile l'allargamento delle formanti e anche la maggiore attenuazione di quelle superiori.

Per mostrare invece lo spostamento delle formanti è stata calcolata l'impedenza d'ingresso di un sistema costituito da due tubi di lunghezza  $L_1$  ed  $L_2$  e raggi  $a_1$  e  $a_2$ : nelle figure da 2 a 6 sono graficati il logaritmo del modulo dell'impedenza d'ingresso per diversi valori delle lunghezze la cui somma è  $L$  e, in ogni figura, per 11 diversi valori dei raggi  $a_k = (1 + k/6)/\sqrt{\pi}$ ,  $k = -5, -4, \dots + 5$ : il raggio del primo tubo decresce dall'alto verso il basso mentre quello del secondo cresce; nel grafico centrale i due raggi sono uguali e quindi il sistema corrisponde a un unico tubo cilindrico di lunghezza  $L = 20$  e raggio  $1/\sqrt{\pi}$ .

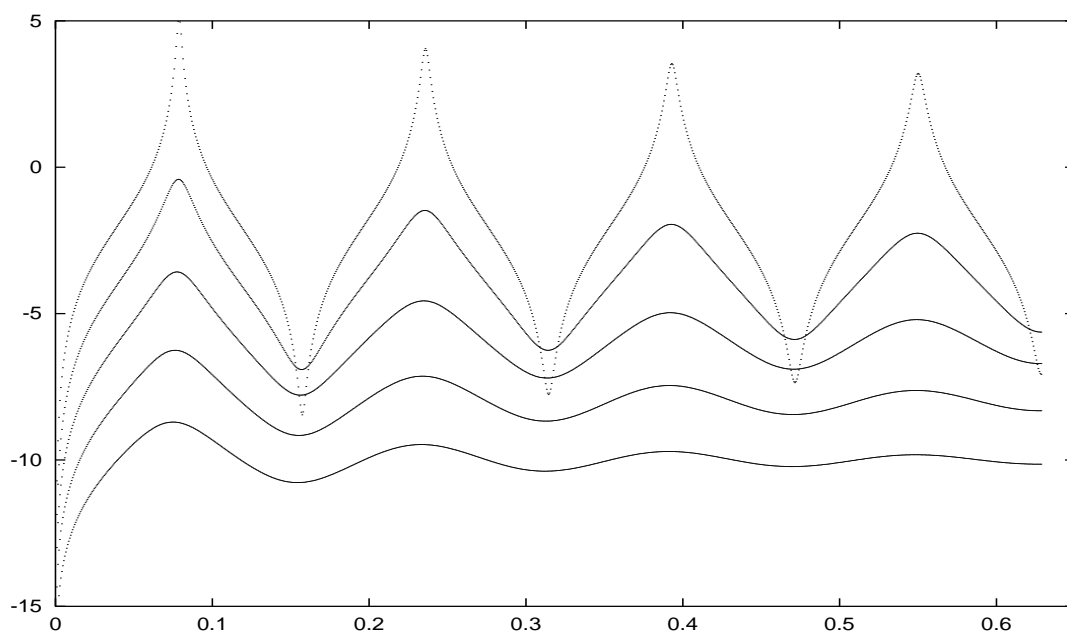


Fig. 12: Le prime quattro formanti di un tubo cilindrico con  $L/v = 20$ ; dall'alto in basso:  $a_k = 0.5k - 0.4$ ;  $k = 1, 2, \dots, 5$ .

Nella Fig.2 i due tubi hanno ugual lunghezza e i minimi dell'impedenza rimangono fissi mentre i massimi si spostano: il primo grafico in alto corrisponde a un tubo molto largo nella prima metà e molto stretto nella seconda e il suo comportamento è simile a quello di un tubo lungo la metà e chiuso all'estremo destro; l'ultimo grafico in basso corrisponde alla situazione opposta e i massimi dell'impedenza (che sono dei doppietti abbastanza stretti) ricordano il comportamento di un tubo chiuso-aperto di metà lunghezza.

Nelle Figg.3 e 4 l'andamento è meno regolare e si osserva lo spostamento anche di alcuni minimi d'impedenza

Nelle Figg.5 e 6 la differenza di lunghezza tra i due tubi è talmente grande che il comportamento, nei due casi estremi, è determinato dal tubo più lungo, mentre quello corto contribuisce a fissare la condizione ad un estremo.

Se con soli due parametri a disposizione (il rapporto tra i raggi e quello tra le lunghezze di due tubi) è possibile controllare abbastanza bene lo spostamento delle formanti è facile immaginare quel che può succedere quando si mettono in serie e/o in parallelo cinque o sei tubi, anche se di questi si possono controllare solo tre raggi equivalenti (labbra, cavo orale e lingua) e l'inserimento o l'esclusione del tratto nasale. Se si dispone di una grande capacità di calcolo è possibile aumentare notevolmente il numero dei tubi per riprodurre in modo più accurato il profilo del tratto vocale, tenendo conto anche dei fenomeni di assorbimento che in questi esempi numerici abbiamo trascurato: in questo caso le formanti oltre a spostarsi acquistano

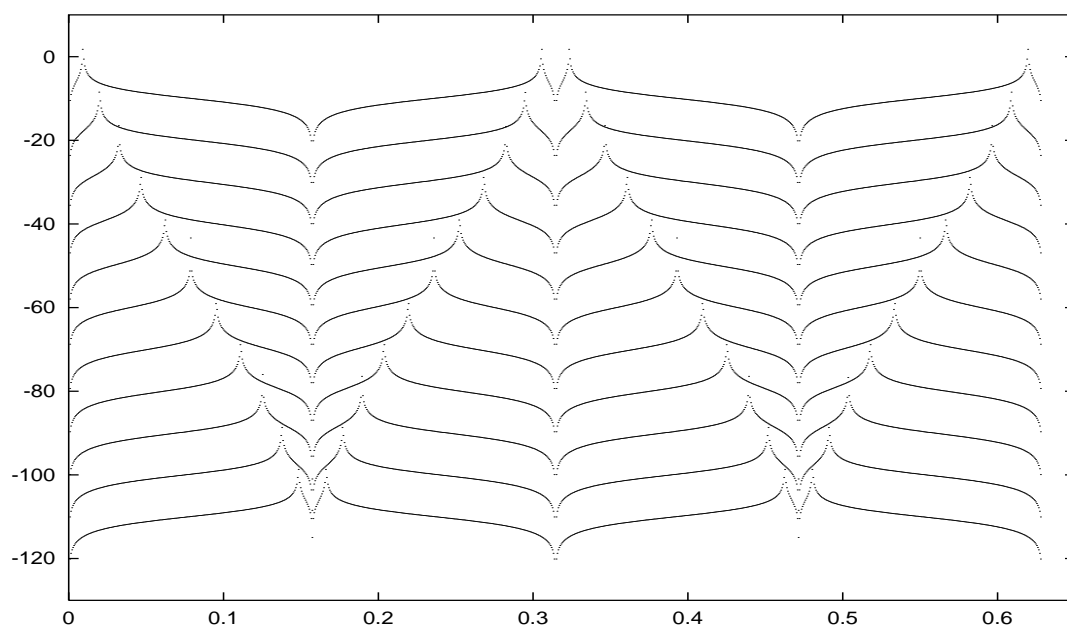


Fig. 13: Logaritmo del modulo dell'impedenza d'ingresso per due tubi in serie;  $L_1 = 10, L_2 = 10$ ;  $a_k = (1 + k/6)/\sqrt{\pi}$ ,  $k = -5, -4, \dots + 4, +5$ .

una larghezza che corrisponde meglio alla situazione reale.

Abbiamo accennato al timbro, che caratterizza una voce, all'interno di questa le vocali e più in generale i suoni sostenuti, ma che cos'è esattamente il timbro, non solo di una voce ma anche di uno strumento? Se si guarda uno spartito musicale non si trovano indicazioni a carattere timbrico se non in forma indiretta, per esempio suggerimenti tecnici sull'esecuzione che influiscono sul timbro: perché? Finora abbiamo caratterizzato i suoni sostenuti secondo l'ampiezza, l'altezza e la distribuzione spettrale: quest'ultima coinvolge l'ampiezza (non la fase) di tutti gli armonici presenti in un suono semplice, che in linea di principio sono infiniti, ma in pratica sono in numero finito, quelli che hanno un'ampiezza che permetta loro di emergere dal rumore di fondo (da un punto di vista oggettivo) o che permetta loro di essere percepiti da un ascoltatore (da un punto di vista soggettivo): si tratta comunque di un insieme di numeri, diciamo dell'ordine della diecina, che caratterizza la posizione di ogni suono in quello che possiamo chiamare "spazio timbrico", al quale possiamo attribuire, oggettivamente, un numero di dimensioni pari al massimo numero di armonici che possono essere effettivamente percepiti. Se accettassimo l'idea che il timbro dipenda da tutti questi parametri dovremmo pretendere che la notazione musicale li contenesse tutti, cosa quasi impossibile.

Non è detto poi che tutte queste dimensioni siano rilevanti dal punto di vista percettivo, e questo per vari motivi, in quanto le alterazioni della distribuzione

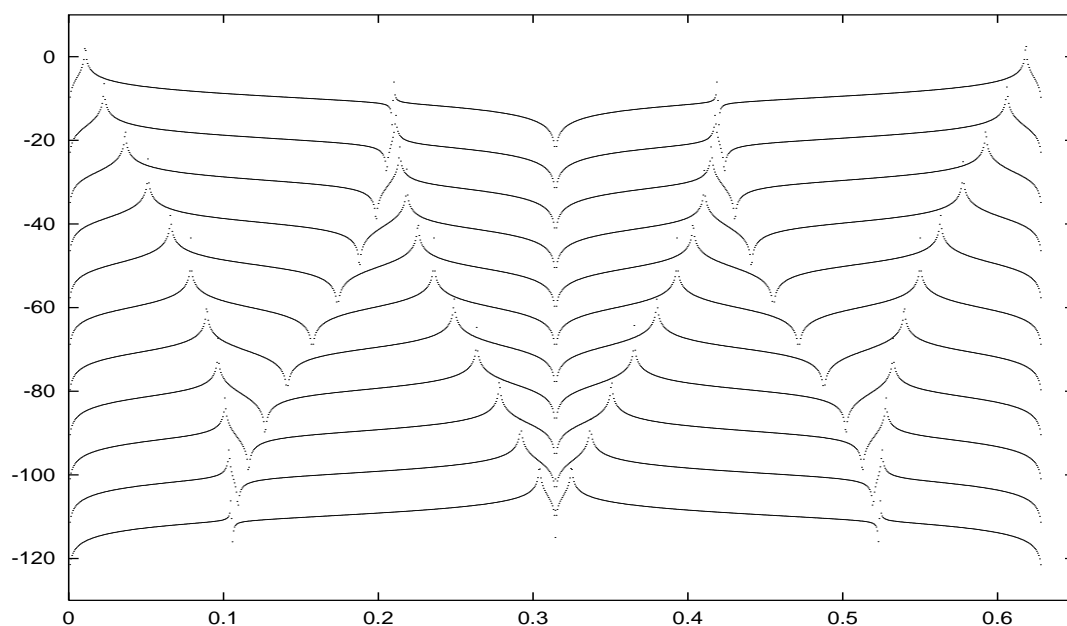


Fig. 14: Logaritmo del modulo dell'impedenza d'ingresso per due tubi in serie;  
 $L_1 = 5, L_2 = 15;$   $a_k = (1 + k/6)/\sqrt{\pi},$   $k = -5, -4, \dots + 4, +5.$

spettrale possono produrre risultati assai diversi: certamente alcune piccole alterazioni non producono alcun cambiamento timbrico degno di nota, mentre alterazioni grandi possono portare ad un cambiamento timbrico significativo, ma anche a un altro fenomeno, che è quello della moltiplicazione del suono originario; se per esempio riduciamo le ampiezze di tutti gli armonici eccettuati il fondamentale e il settimo, ad un certo livello di riduzione si cessa di percepire un unico suono, per esempio LA1, ma si sentono LA1 e SOL3, nettamente distinti.

Diciamo allora che, assegnata la frequenza, l'ampiezza del fondamentale e l'ascoltatore, è possibile definire come spazio timbrico individuale l'insieme di tutte le distribuzioni spettrali che danno luogo alla percezione di un singolo suono: all'interno di questo spazio si possono individuare dei sottospazi e/o delle regioni caratterizzate da un particolare timbro. Ho tenuto a precisare la rilevanza della frequenza, dell'ampiezza e dell'ascoltatore perché è da questi che dipende la dimensionalità dello spazio timbrico individuale: infatti quanto più bassa è la frequenza del fondamentale e maggiore la sua ampiezza, tanto più grande è il numero degli armonici che possono venire percepiti, e questo dipende anche dall'ascoltatore e dalle sue condizioni.

Ma è vero che il timbro dipende soltanto dalla distribuzione spettrale? E da che cos'altro potrebbe dipendere? Se ricordiamo che il nostro orecchio effettua una specie di Trasformata di Fourier a finestra mobile, dobbiamo tener presente che questa viene aggiornata almeno una decina di volte al secondo, e quindi solo per suoni stazionari abbastanza prolungati nel tempo si può parlare di una sola distribuzione spettrale,

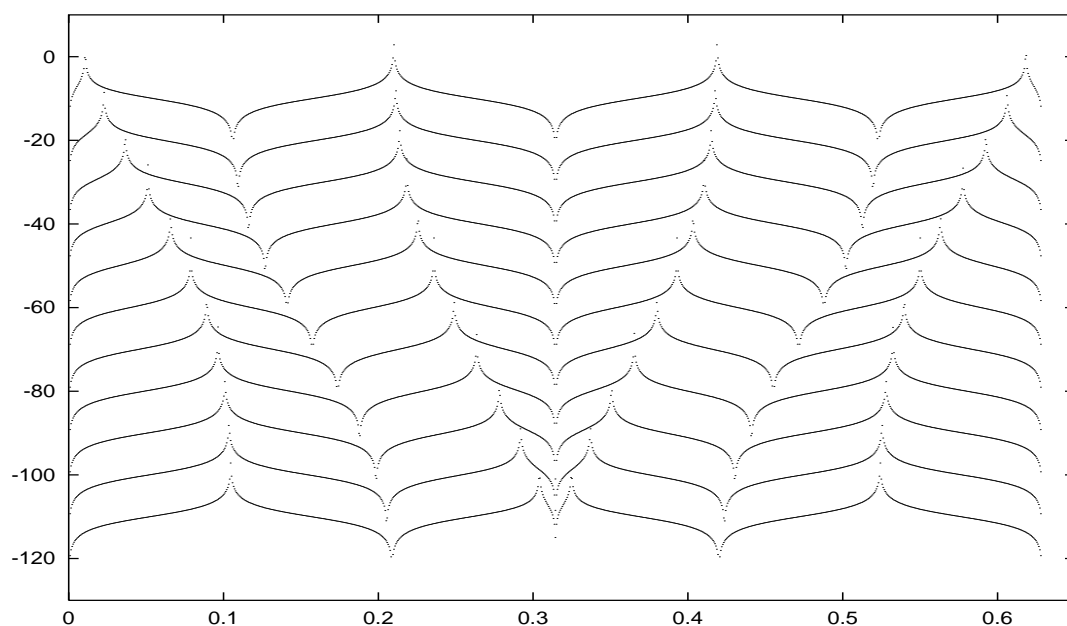


Fig. 15: Logaritmo del modulo dell'impedenza d'ingresso per due tubi in serie;  
 $L_1 = 15, L_2 = 5$ ;  $a_k = (1 + k/6)/\sqrt{\pi}$ ,  $k = -5, -4, \dots + 4, +5$ .

mentre per il transiente d'attacco e per l'eventuale decadimento occorre tener conto anche della sua evoluzione temporale, che è rapida soprattutto nella fase iniziale, ed è probabilmente rilevante nella discriminazione timbrica. Per rendersene conto si possono fare numerosi esperimenti: ascoltare suoni prodotti da strumenti timbricamente diversi dopo averli privati del transiente d'attacco, ascoltare suoni vocalici trattati nello stesso modo e prolungati artificialmente, ascoltare suoni invertiti temporalmente, soprattutto quelli prodotti per pizzico o percussione di una corda. In tutti questi casi riesce molto più difficile discriminare sia i timbri strumentali che le vocali, segno dunque che l'aspetto temporale è forse più rilevante della distribuzione spettrale.

Proviamo a esaminare il problema dal punto di vista dello strumento di misura, cioè l'apparato uditivo inteso nel suo complesso, dall'orecchio esterno alla corteccia uditiva: la sorprendente capacità di separare suoni monocromatici con frequenze che differiscono anche solo del percento ci dice che l'apparato (o una sua sezione) è un ottimo analizzatore a banda stretta, che richiede però tempi relativamente lunghi, a causa del principio d'indeterminazione tempo-frequenza; d'altra parte si sa che per il riconoscimento vocalico (e probabilmente anche timbrico) basta un tempo molto minore, quello necessario per lo sviluppo del transiente d'attacco, al quale corrisponde però un'analisi a banda larga, che non può essere effettuata dalla stessa sezione dell'apparato. Questo non deve sorprendere, visto che una situazione analoga si presenta in campo visivo, dove le forme, i colori, le distanze, sono valu-

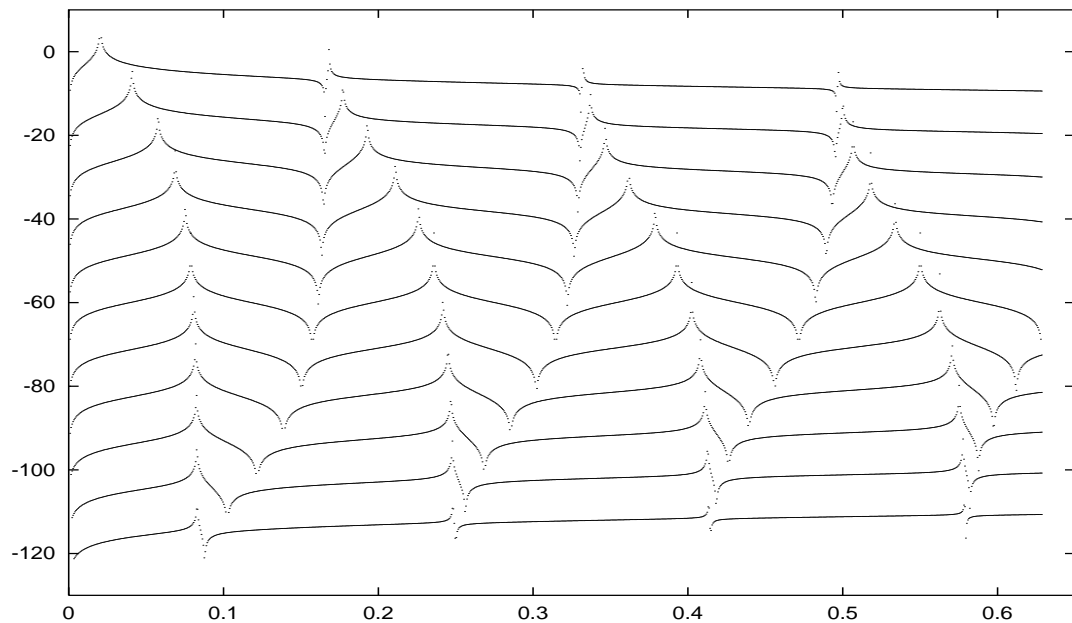


Fig. 16: Logaritmo del modulo dell'impedenza d'ingresso per due tubi in serie;  
 $L_1 = 1, L_2 = 19;$       $a_k = (1 + k/6)/\sqrt{\pi}, \quad k = -5, -4, \dots + 4, +5$

tate in seguito ad elaborazioni effettuate in zone diverse della corteccia visiva, pur utilizzando la stessa informazione primaria, cioè l'eccitazione dei recettori retinici. Anche l'apparato uditivo, oltre alle informazioni di carattere spettrale e temporale (in competizione l'una con l'altra), utilizza la duplicità delle orecchie per spazializzare il suono, cioè per collocare, se non in una posizione, almeno in una direzione, la sorgente del suono.

L'elaborazione delle informazioni uditive ha luogo in entrambi gli emisferi cerebrali, con una suddivisione di compiti: quello destro elabora prevalentemente il contenuto spettrale, quello sinistro quello temporale. D'altra parte le due orecchie conducono l'informazione a entrambi gli emisferi, i quali verosimilmente la utilizzano tutta, sia pure con una preferenza di ciascuno per una parte: è significativo il fatto che le due parti corrispondano agli assi orizzontale e verticale del sonogramma, e che il linguaggio e la musica favoriscano la comunicazione tra i due emisferi.

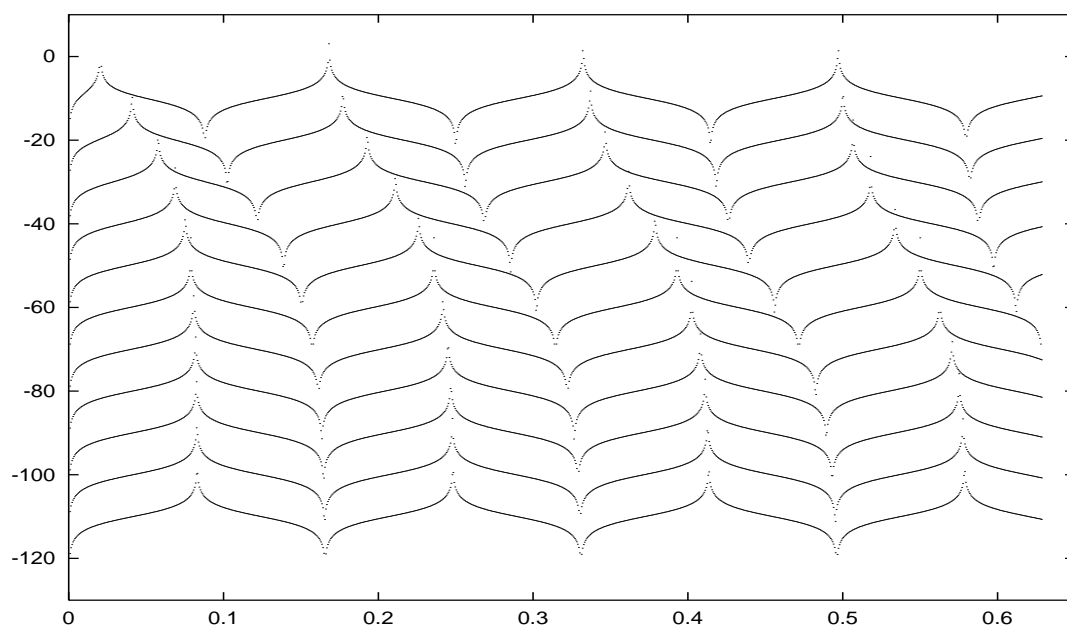


Fig. 17: Logaritmo del modulo dell'impedenza d'ingresso per due tubi in serie;  
 $L_1 = 19, L_2 = 1;$   $a_k = (1 + k/6)/\sqrt{\pi},$   $k = -5, -4, \dots + 4, +5.$