

## Capitolo 6

# Effetti nel dominio spazio-temporale

*Federico Fontana*

Copyright ©1999 by Federico Fontana. All rights reserved.

### 6.1 Introduzione

In questo Capitolo vengono analizzati quegli effetti che simulano o che traggono ispirazione da alcuni fenomeni tipicamente avvertibili in normali condizioni d'ascolto, come il riverbero, l'eco, la colorazione del timbro causata dalle risonanze ambientali, la percezione spaziale di una sorgente sonora. Poichè detti fenomeni sono da mettersi in relazione a modificazioni del segnale acustico tipicamente riconducibili a eventi che avvengono nel dominio del tempo e dello spazio, come ad esempio la riflessione delle onde, gli effetti qui presentati sono accomunati da tecniche progettuali decisamente orientate al trattamento del suono come segnale temporale.

Questo approccio è tanto più motivato se si pensa che, in origine, l'assenza di metodi rigorosi di analisi del segnale costringeva i progettisti a mettere a punto per via perlopiù empirica i modelli, dai quali non di rado sono ugualmente scaturite delle apparecchiature di successo, gradite dai musicisti, dagli ingegneri del suono e dal pubblico al punto di segnare dei traguardi definitivi nella storia degli effetti. È principalmente per questo motivo che, quantunque i modelli si siano molto evoluti, soprattutto con l'avvento dell'elaborazione in tempo reale del segnale numerico e dei *Digital Signal Processor* (DSP), il successo di un effetto è ancora oggi principalmente determinato dalle capacità del progettista, che non di rado si affida alla propria intuizione, e soprattutto alla messa a punto definitiva del sistema sulla base dei pareri dell'ascoltatore.

Nella prima parte si presentano le versioni digitali di una famiglia di effetti popolari ai musicisti, tra cui il *flanger* e il *chorus*, tutti accomunati dal modello che sta alla base delle diverse realizzazioni. Nella seconda parte si affronterà il problema della progettazione di un riverbero. Nella terza parte, infine, si introdurrà il problema della spazializzazione e della binauralizzazione di un segnale musicale, proponendo un modello in grado di svolgere questo compito.

Si insiste sul concetto che le tecniche proposte non possono in alcun modo ritenersi esaustive, costituendo esse solamente una base dalla quale il progettista parte per arrivare infine, grazie alla propria esperienza e a una paziente messa a punto dei parametri del sistema, a un prodotto che possa fregiarsi del nome di *effetto*. Nondimeno, le realizzazioni proposte come esempi, essendo il risultato di tale lavoro, sono da considerarsi a tutti gli effetti come applicazioni efficacemente funzionanti.

## 6.2 Effetti tradizionali

La prerogativa principale degli effetti tradizionali è quella di presentare alla loro uscita il segnale musicale d'ingresso, più un certo numero di ripetizioni di questo. Ciò è sicuramente evidente in un *echo*, ma è vero anche per un *chorus* o un *flanger*. A seconda del tempo che separa le ripetizioni, o *echi*, le leggi della psicoacustica determineranno come viene percepito l'effetto.

Una desiderabile e fondamentale proprietà del sistema progettato è quella di poter modulare il tempo di ritardo, ovvero la separazione temporale tra echi successivi. Accanto a questa prerogativa ve ne sono altre, le quali nel complesso contribuiscono a definire la caratteristica del tipo di effetto.

Il cuore del modello, come vedremo, è costituito dalla *linea di ritardo modulata*, con la quale si può realizzare anche un *vibrato* o un *doubling*. Detto dispositivo è un filtro tempo variante e non si presta a una semplice analisi, anche se la comprensione di alcuni suoi aspetti è fondamentale nell'ambito della trattazione.

### 6.2.1 Un antenato elettromeccanico: Leslie

A titolo introduttivo, si richiama la modalità di funzionamento del Leslie, il quale in un certo senso è l'analogo elettromeccanico della linea di ritardo modulata<sup>1</sup>. Esso consiste (vedi Figura 6.1) in una

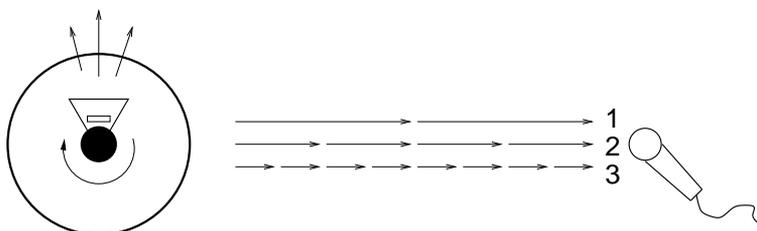


Figura 6.1: Schema di funzionamento del Leslie. La componente della velocità dell'onda acustica calcolata lungo l'asse che collega l'asse di rotazione al microfono è proporzionale alla lunghezza delle frecce, in corrispondenza delle rispettive posizioni dell'altoparlante.

cassa di forma circolare, all'interno della quale gira a una velocità prestabilita un altoparlante, non troppo direttivo ed eccentrico rispetto all'asse di rotazione.

1. L'ascoltatore ode dapprima il suono emesso dall'altoparlante che, ruotando eccentricamente, si sta avvicinando a lui (onda contrassegnata con 1);
2. nel momento in cui la componente della velocità misurata lungo l'asse che collega l'ascoltatore al diffusore è nulla, ode l'onda contrassegnata con 2;
3. infine, quando il diffusore si sta allontanando, ode l'onda contrassegnata con 3.

L'effetto Doppler assicura la percezione di un segnale modulato, il cui periodo di modulazione è determinato dalla velocità di rotazione, e la cui profondità di modulazione deriva dal rapporto di eccentricità del diffusore rispetto all'asse di rotazione.

Il successo ottenuto dal Leslie è sicuramente dovuto alla timbrica unica impressa al suono, e al senso di piacevolezza suscitato dalla sensazione del vibrato.

<sup>1</sup>Nonostante ciò, non è consigliabile invitare il possessore di un Leslie a disfarsi del proprio gioiello per sostituirlo con un DSP.

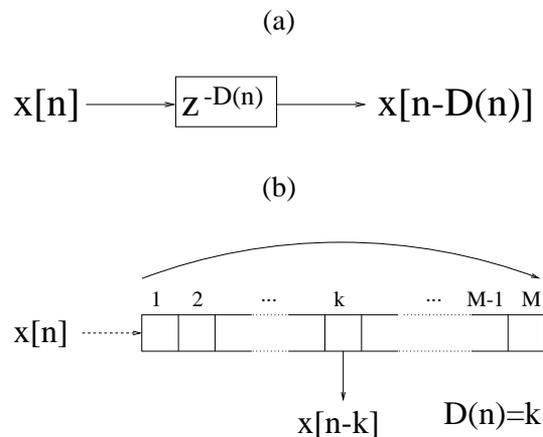


Figura 6.2: (a) Linea di ritardo modulata e (b) sua realizzazione mediante coda circolare.

### 6.2.2 La linea di ritardo modulata

La possibilità di definire una linea di ritardo modulata in un sistema a tempo discreto appare, formalmente, piuttosto semplice. Basta infatti implementare lo schema di Figura 6.2(a), dove il ritardo è regolato dalla *funzione di modulazione*  $D$ , per avere a tutti gli effetti creato l'oggetto che fa al caso nostro. La Figura 6.2(b) evidenzia un possibile schema realizzativo, che sfrutta una coda circolare in cui, all'istante  $n$ -esimo, la testa è individuata dalla freccia tratteggiata, mentre l'uscita è posta in corrispondenza della locazione  $k$ -esima, da cui esce la freccia a linea continua.

La dimensione della coda individua gli estremi oltre i quali la funzione  $D$  non può assumere valori; se dunque la coda ha dimensione uguale a  $M$ ,  $D(n)$  dovrà rispettare la condizione

$$0 \leq D(n) \leq M. \quad (6.1)$$

Più complessa appare l'analisi rigorosa dal punto di vista della teoria del segnale discreto. In ogni caso l'oggetto, per come è stato definito, sarebbe di qualche utilità solamente all'interno di un sistema che lavori a frequenze di campionamento ben più elevate di quelle richieste dalla banda audio. Infatti una funzione di modulazione, il cui codominio sia contenuto negli istanti di campionamento di un segnale campionato in banda audio, non è certamente in grado di produrre una modulazione convincente. La cosa appare ancora più evidente dagli esempi seguenti.

◦ Esempio: cambio di pitch. Si ha:

$$D(n) = (1 - \text{pcr})n$$

dove  $\text{pcr}$  (*pitch change ratio*) è il rapporto di variazione del pitch tra il segnale processato dalla linea di ritardo modulata e il segnale originale. Per  $\text{pcr} = 2$ , ad esempio, si ha un raddoppio delle altezze, infatti l'uscita dalla linea di ritardo modulata vale

$$x[n - D(n)] = x[\text{pcr} \cdot n].$$

Ovviamente, per mantenere indefinitamente il cambio di pitch occorrerebbe una coda di dimensione illimitata. Altrimenti il tempo massimo per cui l'effetto può essere mantenuto vale

$$\left\lfloor \frac{M}{1 - \text{pcr}} \right\rfloor T$$

secondi, con  $T$  periodo di campionamento del sistema.

◦ Esempio: vibrato. È

$$D(n) = \frac{M}{2} \sin(2\pi f_0 n) + \frac{M}{2}$$

e dunque

$$x[n - D(n)] = x\left[n - \frac{M}{2} - \frac{M}{2} \sin(2\pi f_0 n)\right].$$

Considerando il segnale  $x$ , secondo l'analisi di Fourier, come la sovrapposizione lineare di componenti sinusoidali, abbiamo che la componente generica posta alla frequenza  $f_x$  viene modulata a una frequenza  $f_m$ , la quale può essere calcolata derivando rispetto al tempo l'argomento relativo alla componente in esame, uguale per esempio  $A_x \sin(2\pi f_x n + \phi_x)$ :

$$f_m = \frac{1}{2\pi} \frac{\partial}{\partial n} \{2\pi f_x [n - \frac{M}{2} - \frac{M}{2} \sin(2\pi f_0 n)] + \phi_x\} = f_x \{1 - 2\pi f_0 \frac{M}{2} \cos(2\pi f_0 n)\}.$$

Dunque il pcr risulta essere uguale a

$$\frac{f_m}{f_x} = 1 - 2\pi f_0 \frac{M}{2} \cos(2\pi f_0 n),$$

Si evince che il pcr non dipende dall'altezza del segnale originale; il suo periodo dipende dalla modulante  $f_0$ , e infine la profondità di modulazione è determinata da  $f_0$  e  $M$ .

Dagli esempi emerge chiaramente il limite imposto dall'uso di una funzione di modulazione "discretizzata" sul periodo di campionamento. Se nel primo caso un tentativo di *pitch change* si può fare imponendo un valore intero per il pcr, nel secondo caso la legge che governa il vibrato salta non appena il prodotto  $(M/2) \sin(2\pi f_0 n)$  sia frazionario per qualche valore di  $n$ .

### 6.2.3 Interpolazione

Se potessimo estendere il codominio di  $D$  a un insieme più ampio, diciamo un certo numero di frazioni dell'unità di campionamento, riusciremmo a istituire delle efficaci funzioni di modulazione altrimenti improbabili. La questione equivale a estendere la  $D$  a una nuova funzione  $D_c$  definita su di un codominio più ampio. In tal caso, si può sempre esprimere un valore assunto da  $D_c$  come la somma del valore di  $D$  immediatamente inferiore, più un numero  $\alpha$ , definito all'interno di un insieme di fissata densità, compreso tra zero e uno:

$$D_c(n) = D(n) + \alpha, \quad 0 \leq \alpha < 1.$$

Detto questo, nel discreto non ha senso pensare di acquisire il campione intermedio  $x[n - D(n) - \alpha]$ . Si ricorre di conseguenza all'*interpolazione*, ovvero al calcolo approssimato del suo valore. Allo scopo, nel nostro sistema sarà presente un interpolatore in grado di produrre un segnale  $y$  il cui valore, secondo qualche criterio, imiti quello del campione intermedio.

Nel seguito vediamo i modelli di interpolazione più usati.

#### Interpolazione lineare

Si pone

$$y[n] = \sum_{k=0}^{L-1} \alpha_k(n) x[n - D(n) - N_k] \quad (6.2)$$

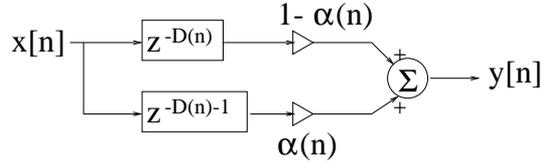


Figura 6.3: Sistema costituito da linea di ritardo modulata e interpolatore lineare.

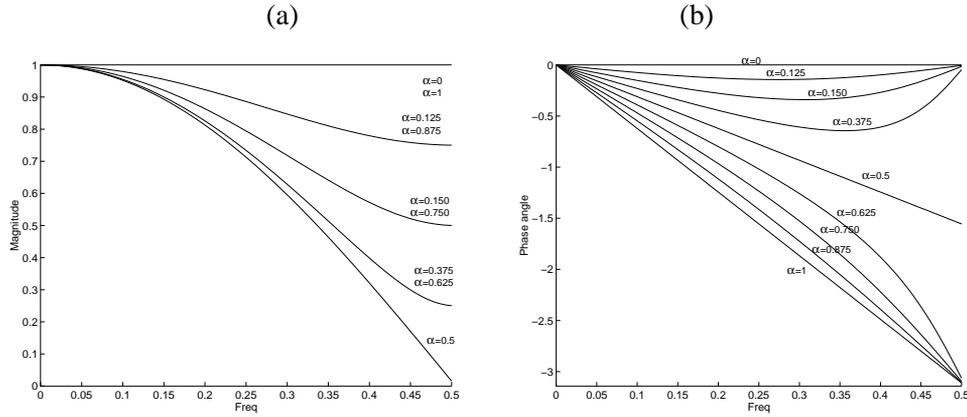


Figura 6.4: (a) moduli della risposta in frequenza e (b) diagrammi di fase del sistema costituito da linea di ritardo modulata più interpolatore lineare.

con

$$0 < \alpha_k(n) < 1 \forall k \text{ e } N_k \in \mathcal{N} \forall k.$$

Il campione intermedio dunque è ottenuto come combinazione lineare di  $L$  campioni. Il caso  $\alpha_k(n) = 1$  avviene in particolare quando  $D_c(n) = D(n) - N_k$ , e in tal caso tutti gli altri coefficienti saranno nulli. Ovviamente deve essere sempre rispettata la condizione

$$0 \leq N_k + D(n) \leq M.$$

Il caso più semplice di interpolazione lineare, molto usato nelle applicazioni tempo reale, si ha per  $L = 2$ ,  $N_0 = 0$  e  $N_1 = 1$ . In tal caso il contributo dei due campioni adiacenti al campione intermedio, distante  $\alpha$  dal campione più recente, viene pesato da coefficienti tra loro complementari:

$$y[n] = \{1 - \alpha(n)\}x[n - D(n)] + \alpha(n)x[n - D(n) - 1].$$

L'interpretazione del modello è piuttosto ovvia e viene lasciata come esercizio.

Il sistema complessivo costituito dalla linea di ritardo modulata più l'interpolatore lineare è schematizzato in Figura 6.3, e si realizza in pratica predisponendo due punti d'uscita consecutivi sulla coda circolare. La Funzione di Trasferimento (FdT) del sistema è

$$\frac{Y(z)}{X(z)} = z^{-D(n)} \{1 - \alpha(n) + \alpha(n)z^{-1}\} \quad (6.3)$$

Da questa, come noto, si possono ricavare modulo e fase della risposta del sistema. Questi sono visualizzati rispettivamente in Figura 6.4(a) e 6.4(b), al variare del parametro  $\alpha$ . Notiamo i seguenti fatti:

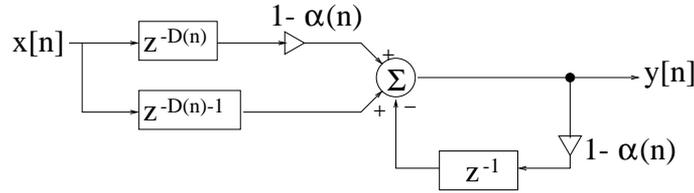


Figura 6.5: Diagrammi di fase del sistema costituito da linea di ritardo modulata più interpolatore allpass.

- l'interpolatore lineare introduce una *distorsione di ampiezza*, attenuando le alte frequenze nei casi non banali  $\alpha = 0$  e  $\alpha = 1$ ;
- l'interpolatore lineare introduce una *modulazione di ampiezza*, e ciò è dovuto alla variazione nel tempo del modulo dello spettro, il quale come detto dipende da  $\alpha(n)$ ;
- l'interpolatore lineare introduce una *distorsione di fase*, in quanto non è un filtro a fase lineare se non nei casi  $\alpha = 0$ ,  $\alpha = 1$  e  $\alpha = 1/2$ . Si noti comunque la linearità della fase in bassa frequenza dove, in più, la pendenza delle curve è proprio uguale a  $\alpha$ ;
- l'interpolatore lineare introduce una *modulazione di fase*, e ciò è dovuto alla variazione nel tempo del diagramma di fase.

Lo studio rigoroso di un filtro a coefficienti tempo varianti non si esaurisce certo in una analisi della risposta al variare dei parametri<sup>2</sup>, e di questo dovremo tenere conto discutendo il prossimo interpolatore. Nel caso dell'interpolatore lineare, comunque, l'analisi fatta è sufficiente a comprendere i limiti del semplice schema presentato.

Un miglioramento delle prestazioni si ottiene aumentando la complessità dell'interpolatore, ovvero aumentando il valore di  $L$ ; questo, a patto di aumentare il costo del sistema. Alternativamente, in alcuni casi si ricorre a una efficiente alternativa, descritta nel seguito.

### Interpolazione allpass

In questo caso il campione in uscita dall'interpolatore viene così calcolato:

$$y[n] = \{1 - \alpha(n)\}x[n - D(n)] + x[n - D(n) - 1] - \{1 - \alpha(n)\}y[n - 1].$$

L'operazione complessiva di modulazione del segnale viene realizzata dallo schema di Figura 6.5, nel quale è evidente la retroazione dell'uscita. La FdT è

$$\frac{Y(z)}{X(z)} = z^{-D(n)} \frac{1 - \alpha(n) + z^{-1}}{1 + \{1 - \alpha(n)\}z^{-1}} \quad (6.4)$$

dalla quale non è difficile ricavare l'unità del modulo della risposta a tutte le frequenze.

Diversamente, il diagramma di fase (Figura 6.6) evidenzia ancora la presenza della distorsione e della modulazione di fase. La linearità in bassa frequenza sussiste ancora, ma la pendenza nell'origine della curva associata a  $\alpha$  non coincide col valore del parametro. Eventualmente, attraverso una

<sup>2</sup>Il lettore attento avrà notato l'aporia presente nella Zeta trasformata (6.3), nella quale in effetti non sono state trasformate le  $\alpha(n)$ .

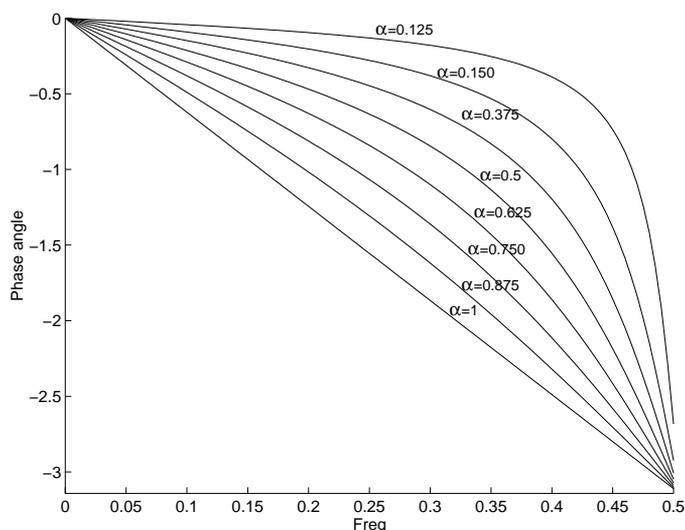


Figura 6.6: Sistema costituito da linea di ritardo modulata e interpolatore allpass.

trasformazione, si può associare al parametro il filtro il cui diagramma di fase abbia identica pendenza nell'origine.

L'interpolazione allpass quindi, nei limiti dell'analisi condotta, risolve le questioni legate alla distorsione e modulazione di ampiezza, irrisolte dal precedente interpolatore.

### Distorsione

Applicando i due diversi interpolatori su un modello per l'alterazione costante del pitch, implementato su un sistema a 16 bit in aritmetica a virgola fissa, che campiona a 44.1 kHz, si può calcolare la *Total Harmonic Distortion plus Noise* (THD+N) dell'uscita effettiva rispetto a un'uscita ideale. I risultati compaiono nel diagramma di Figura 6.7, dove la THD+N viene calcolata al variare del pcr imposto.

L'interpolatore lineare (a) mantiene le proprie prestazioni pressochè costanti al variare del pcr. Questo risultato è in accordo con l'analisi fatta in precedenza, che nel complesso afferma la costanza della distorsione armonica totale.

L'interpolatore allpass, viceversa, mostra una spiccata dipendenza della THD+N dal pcr. Questo fatto evidenzia i limiti dell'analisi condotta sull'interpolatore, la quale in effetti resta valida solamente per *piccole variazioni nel tempo della funzione di modulazione*<sup>3</sup>, cioè di  $\alpha$ .

Analoghe considerazioni emergono applicando i due diversi interpolatori su un modello per la modulazione del pitch. In più, in questo caso la THD+N varia nel tempo, analogamente al pcr.

### 6.2.4 Effetti

Quanto illustrato in precedenza ha posto le premesse per comprendere lo schema generale, mostrato in Figura 6.8, sul quale si basano gli effetti tradizionali realizzati nel dominio discreto. In esso riconosciamo

<sup>3</sup>Basti pensare che, in un filtro tempo variante, l'evoluzione dello stato dipende a sua volta dalla variabile temporale; ciò fa sì che, nel nostro caso d'esempio, non sia nemmeno assicurata la caratteristica allpass dell'interpolatore. Piccole variazioni di  $\alpha$  in effetti fanno propendere per una maggiore credibilità dell'analisi semplificata fatta in precedenza.

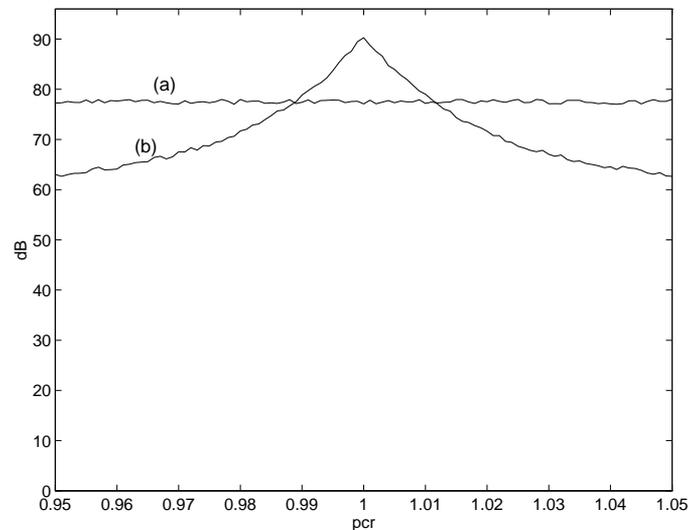


Figura 6.7: THD+N in un modello per l'alterazione del pitch usando l'interpolazione lineare (a) e l'interpolazione allpass (b).

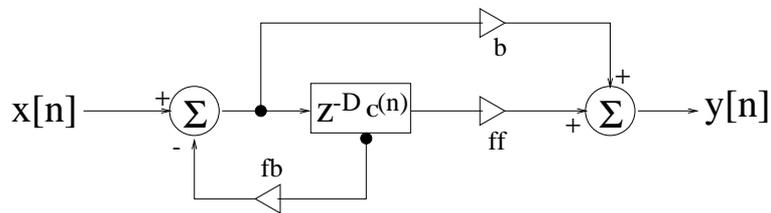


Figura 6.8: Schema generale di effetto tradizionale digitale.

- il complesso della linea di ritardo modulata più l'interpolatore, riassunto nel blocco di FdT uguale a  $z^{-D_c(n)}$ ;
- un *feedback*, controllato dal coefficiente fb e scelto in corrispondenza di un fissato punto della coda, dal quale si riconviola all'ingresso un segnale avente Zeta trasformata uguale a

$$\frac{fbz^{-D}}{1 + fbz^{-D}}X(z),$$

con D ritardo scelto per il feedback, tale che  $1 \leq D \leq M$  e  $|fb| < 1$ . In effetti retroazionare il segnale modulato, ovvero modularlo ricorsivamente, avrebbe come risultato l'impossibilità di controllare la modulazione stessa;

- l'uscita modulata, pesata dal coefficiente di *feedforward* ff;
- un segnale non modulato, che viene miscelato a quello modulato ed è controllato mediante il coefficiente di *blend* b.

La scelta dei parametri messi a disposizione dallo schema di Figura 6.8 dà già la possibilità di selezionare un discreto numero di effetti diversi. Requisiti di economia fisseranno di fatto la scelta dell'interpolatore e di tutte le altre componenti critiche.

**Vibrato**

$$b = 0, ff = 1, fb = 0, 0 \leq MT \leq 5 \text{ ms}$$

Eliminando feedback e blending, acquisendo dunque all'uscita il solo segnale modulato, si ha un effetto vibrato. La linea di ritardo sarà dimensionata per gestire un ritardo approssimativamente inferiore a 5 ms. Un ritardo minimo, anche inferiore a 1 ms, dà i migliori risultati.

**Flanger**

$$b = \sqrt{2}, ff = 1/\sqrt{2}, fb = -1/\sqrt{2}, 0 \leq MT \leq 10 \text{ ms}$$

Scopo del flanger è sovrapporre al segnale d'origine un segnale ritardato dinamicamente, il cui ritardo resti all'interno del periodo d'integrazione dell'orecchio. Allo scopo, una linea di ritardo di 1 ms va bene per molte applicazioni.

Una caratteristica apprezzata nel segnale risultante è il frastagliamento dello spettro. I coefficienti proposti massimizzano l'effetto *comb* spettrale, ovvero la sagomatura dello spettro a denti di pettine. Per quanto riguarda la scelta dell'interpolatore, essa cadrà sull'allpass, in quanto l'attenuazione di ampiezza imposta dall'interpolatore lineare va a discapito della presenza di denti ben definiti alle frequenze più alte.

**Chorus standard**

$$b = 1, ff = 1/\sqrt{2}, fb = 0, 1 \leq MT \leq 30 \text{ ms}$$

Come dice il nome, il chorus è stato pensato per sovrapporre al suono originale una sua replica ritardata (5 ms vanno generalmente bene), in modo tale da generare un effetto simile a quello di un coro aggiunto al segnale musicale.

In questo caso i denti nello spettro sono indesiderabili, sicché una semplice soluzione consiste nel tagliare il feedback.

La trasparenza è un parametro molto apprezzato da chi utilizza assiduamente il chorus, come i chitarristi. L'interpolatore allpass, per le note non linearità che induce, non si presta dunque ad essere applicato nel chorus, e tuttavia l'interpolatore lineare presenta la tipica attenuazione. Una soluzione raffinata è quella adottata nei *white chorus*, dove si retroaziona il segnale, mantenendo il sistema globale il più possibilmente allpass, cioè imponendo  $fb = b$ . Il sistema in tal modo presenta un'uscita più ricca in alte frequenze, pur avendo rinunciato all'interpolatore allpass. La presenza di picchi residui nello spettro viene ulteriormente attenuata mantenendo  $ff \neq b$ . Una scelta per i coefficienti può essere

$$b = 1/\sqrt{2}, ff = 1, fb = 1/\sqrt{2}.$$

**Doubling**

$$b = 1/\sqrt{2}, ff = 1/\sqrt{2}, fb = 0, 10 \leq MT \leq 100 \text{ ms}$$

Un effetto usato soprattutto dai cantanti è quello di raddoppiare la traccia del cantato, rieseguendo la parte che va così a sovrapporsi quella già esistente. Qui il margine di ritardo può variare discretamente (20 ms tipicamente vanno bene): una modulazione abbastanza casuale, come quella causata dal doppio cantato, è auspicabile.

**Echo**

$$b = 1, ff \leq 1, fb < 1, MT \geq 50\text{ms}$$

L'eco si distingue per la lunghezza della linea di ritardo, che dovrebbe assicurare un ritardo di almeno 80 ms, tempo al di sotto del quale non è assicurata la percezione distinta di due suoni identici. I coefficienti vengono perlopiù tarati in base alla timbrica richiesta per il tipo di eco.

**6.3 Riverbero**

Chi non ha mai visitato una camera anecoica, non può immaginare quanta importanza abbia il riverbero nella percezione degli eventi sonori negli ambienti chiusi. Ci si trovi in una piccola stanza o in una sala da concerto, l'insieme delle risonanze e degli echi che accompagnano il suono puro influiscono in modo fondamentale sulla nostra percezione del timbro e dell'ambiente che ci circonda. Test durante i quali a un ascoltatore, collocato al buio in un piccolo ambiente, veniva fatto ascoltare un evento sonoro caratterizzato da un riverbero da sala da concerto, hanno dimostrato che tale effetto risultava desiderabile e ben riconoscibile fintantochè all'ascoltatore non veniva palesato il fatto di trovarsi all'interno di un piccolo ambiente; a quel punto egli avvertiva quella sensazione chiamata *straniamento*, secondo la quale l'ascoltatore si trova in una situazione sperimentale "perturbata", non più riconoscibile.

Appurata l'importanza di caratterizzare l'evento sonoro anche mediante un corretto riverbero, bisogna considerare le opposte esigenze di chi affronta il problema della registrazione musicale. Una registrazione condotta all'interno di un ambiente riverberante risulta pessima nella maggioranza dei casi, e questo non solo perchè il tipo di riverbero, legato alle caratteristiche dell'ambiente, quasi certamente non sarà quello desiderato dal musicista<sup>4</sup>, ma anche perchè, non appena i musicisti siano più di uno, fenomeni di *cross-talk* tra uno strumento e l'altro andranno a rimescolare le tracce registrate. Questo obbliga a rendere una sala di registrazione la più anecoica possibile, e a minimizzare l'influenza sulla singola traccia di qualunque segnale esterno.

Da queste opposte considerazioni appare evidente l'esigenza di disporre, in sede di elaborazione del segnale musicale, di sistemi in grado di riverberare un suono puro nel modo più *naturale* possibile.

Nel proseguio vedremo dapprima brevemente il riverbero come fenomeno fisico, e alcuni effetti percettivi che esso induce. Poi si spiegheranno le problematiche esistenti nella scelta di un modello di rappresentazione del riverbero, e le risposte fornite dai progettisti. Infine si proporranno alcuni esempi, in ordine di difficoltà crescente. Tale ordine coincide con quello storico, per quelle considerazioni già fatte nell'Introduzione, sulle quali non è il caso di tornare.

**6.3.1 Considerazioni fisiche sul riverbero**

Si supponga di sollecitare un ambiente riverberante con un segnale impulsivo. Contemporaneamente, si acquisisca il segnale  $h$  presente in corrispondenza di un punto nell'ambiente — conseguenza della sollecitazione prodotta — ad esempio mediante un microfono. Assumendo la linearità della catena di misura, è lecito considerare il segnale acquisito come una *risposta impulsiva* dell'ambiente. La risposta dipende dalle posizioni relative della sorgente di segnale e del ricevitore, ed esistono tecniche precise per l'acquisizione e il calcolo di questo tipo di risposte, delle quali non ci occupiamo.

Un tipico esempio dell'andamento nel tempo di una risposta impulsiva ambientale è illustrato in Figura 6.9. L'esempio si riferisce a un piccolo ambiente.

Analizziamo il segnale nel dettaglio:

<sup>4</sup>Per un certo tipo di musicista è d'obbligo togliere il quasi.

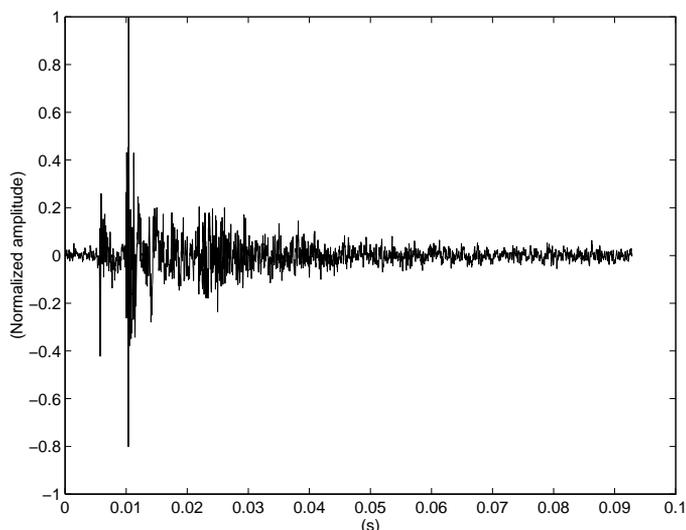


Figura 6.9: Andamento nel tempo di una risposta impulsiva ambientale.

- dopo circa 7 ms, prima dei quali viene solamente registrato del rumore di fondo, arriva al punto di ricezione il *segnale diretto* (*direct signal*);
- successivamente cominciano ad arrivare le *prime riflessioni* (*early reflections*), provenienti dalle pareti riflettenti dell'ambiente, non tutte di ampiezza necessariamente minore a quella dell'impulso diretto<sup>5</sup>;
- con l'avanzare del tempo, alle prime riflessioni si uniscono le riflessioni di ordine superiore o *riflessioni successive* (*late reflections*), che perdurano fintantochè le pareti dell'ambiente e l'attenuazione dell'aria non riducono la potenza del segnale a un valore trascurabile. Si usa anche dire che, cessato l'effetto delle prime riflessioni, il campo acustico diviene *diffuso*.

In generale la legge di decadimento della pressione acustica e dell'energia di un segnale riverberante è di tipo esponenziale. In effetti, a questa legge si attiene l'involuppo del segnale di Figura 6.9. In questo contesto, una misura particolarmente significativa delle caratteristiche riverberanti di un ambiente è il *tempo di riverbero ai 60 dB*, denotato con  $RT_{60}$ , ovvero il tempo in secondi che trascorre affinché l'involuppo dell'energia della risposta impulsiva ambientale si riduca di 60 dB.

Per calcolare con precisione il tempo di riverbero ai 60 dB non è necessario mediare un elevato numero di involuppi, dedotti rispettivamente da misure eseguite spostando la sorgente e/o il ricevitore in più punti. È stato dimostrato che la curva di decadimento energetico  $RT(t)$  del segnale si ottiene calcolando la funzione

$$RT(t) = \int_t^{\infty} h^2(\tau) d\tau, \quad (6.5)$$

al variare del tempo  $t$ .

Si noti fin d'ora l'elevato frastagliamento del segnale di Figura 6.9. Un modello statistico per le riflessioni dimostra che il *numero di echi*  $N_t$  misurati dopo un tempo  $t$  trascorso dall'emissione di un segnale, che si generano per riflessione all'interno di un ambiente chiuso di volume  $V$  dove il segnale

---

<sup>5</sup>Perchè ?

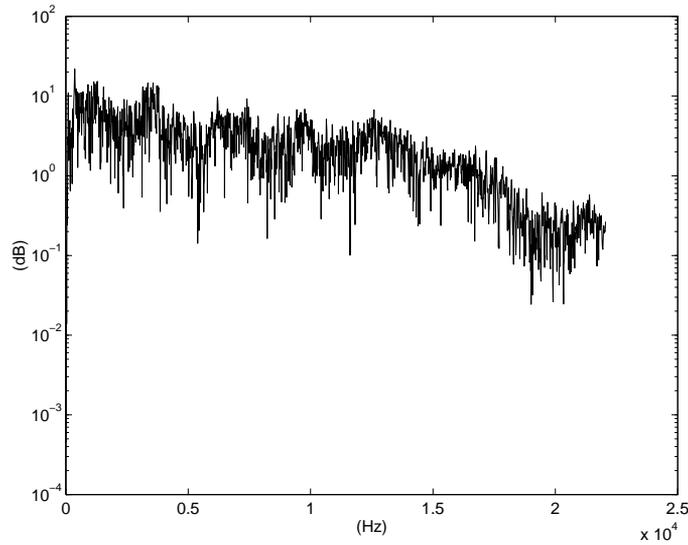


Figura 6.10: Modulo della trasformata di Fourier della risposta impulsiva ambientale di Figura 6.9.

si propaga a una velocità  $c$ , è uguale a

$$N_t = \frac{4\pi c^3}{3V} t^3. \quad (6.6)$$

Dalla formula si vede che il numero di echi cresce nel tempo secondo una legge cubica.

L'analisi in frequenza della risposta impulsiva ambientale fornisce delle informazioni altrettanto interessanti. La Figura 6.10 mostra il modulo della trasformata di Fourier del segnale di Figura 6.9. In essa si notano in particolare dei picchi in corrispondenza di determinate frequenze. A ciascun picco è associato un *modo* di risonanza.

La teoria afferma che i modi vanno addensandosi all'aumentare della frequenza. Essi restano distinguibili nello spettro, e dunque entro certi limiti anche a livello percettivo, fino a una certa frequenza che varia con i parametri ambientali. Al di sopra di questa frequenza i modi tendono progressivamente a sovrapporsi, rendendosi di fatto indistinguibili. Si dimostra, sempre sulla base di un modello statistico, che il numero di modi normali  $N_f$  presenti fino alla frequenza  $f$  è uguale a

$$N_f = \frac{4\pi V}{3c^3} f^3. \quad (6.7)$$

Anche la (6.7) dunque, duale rispetto alla (6.6), sancisce la progressione cubica del numero dei modi con la frequenza. In ogni caso sia la (6.7) che la (6.6), avendo validità statistica, non sono utili per conteggiare le prime riflessioni o, dualmente, i modi in bassa frequenza.

Infine, vale la pena ricordare la formula che permette di stimare la *separazione media*  $\Delta f_{\max}$  tra i picchi spettrali, misurata (in Hz) in corrispondenza della banda dove la densità dei modi giustifica un loro conteggio su base statistica:

$$\Delta f_{\max} \approx \frac{4}{RT_{60}}. \quad (6.8)$$

### 6.3.2 Percezione del riverbero

Il riverbero contribuisce in modo fondamentale alla caratterizzazione di un evento sonoro. Assumiamo fin d'ora di tralasciare quegli aspetti capaci di influenzare la nostra percezione spaziale, che per un

discorso di semplicità consequenziale vogliamo accorpate tutti nella Sezione successiva, nonostante su questa scelta si possano avanzare fondate obiezioni.

Le caratteristiche dell'evento sonoro che maggiormente risentono del tipo di riverbero sono:

- il *volume* del suono diretto;
- il *colore* del suono;
- il *timbro*;
- il *tappeto acustico* che si sovrappone al suono diretto, ovvero la percezione di uno “sfondo” al suono diretto;
- l'insieme degli *echi* che vanno ad accodarsi al suono diretto.

La letteratura sull'argomento è ampia e qualche volta in disaccordo sui risultati. Generalmente, si assume di suddividere la percezione temporale di un evento sonoro basandosi sulla *costante di integrazione dell'orecchio*, fissata in 80 ms. In base a questo modello, valgono le seguenti considerazioni di principio:

1. a partire dall'istante di percezione del segnale diretto, tutti gli echi che giungono all'orecchio entro 80 ms vanno a caratterizzare volume, colore e timbro del suono;
2. le caratteristiche del tappeto acustico dipendono in larga parte dagli echi attenuati che giungono dopo 80 ms. Di fatto, durante l'ascolto spesso il tappeto è mascherato dal segnale diretto, e viene perlopiù valutato durante le pause;
3. ripetizioni del segnale diretto, che giungono dopo un tempo superiore alla costante di integrazione dell'orecchio, vengono percepite come eventi distinti, o echi veri e propri.

### 6.3.3 Modelli

La fase più difficile nel processo di messa a punto di un riverbero è costituita dalla scelta di un modello adeguato. I tentativi di conciliare i parametri fisici e quelli percettivi non sono generalmente andati a buon fine, sicché una scelta decisiva va compiuta già nella fase di approccio al modello. Nondimeno, sono stati scoperti alcuni fondamentali legami tra le caratteristiche spettrali del segnale e la percezione del riverbero, dei quali tratteremo nel seguito.

#### L'approccio “fisico”

L'approccio fisico ha il vantaggio di mettere a punto dei modelli dotati di parametri fisici accessibili, come dimensioni e forma dell'ambiente, riflessività delle pareti e attenuazione ambientale, posizioni della sorgente e del punto di ascolto e così via.

La soluzione più semplice consiste nel memorizzare a bordo del riverbero tutta l'informazione relativa a un insieme di  $N$  risposte impulsive ambientali campionate  $h_1, \dots, h_N$ , e di calcolare il segnale riverberato  $y$  come convoluzione discreta tra il segnale musicale  $x$  e una risposta ambientale  $h_i$  scelta all'interno di un database:

$$y(n) = h_i * x(n)$$

È ovvio che questa scelta pone dei seri vincoli sulla possibilità di implementare l'effetto in tempo reale, causa l'elevata quantità di memoria e di potenza di calcolo richieste da un simile modello.

In alternativa, sono stati proposti diversi modelli di propagazione del segnale all'interno di un ambiente, con l'intento di semplificarne la descrizione al fine di poter avere un guadagno in termini di efficienza. Di questi vanno ricordati:

- l'*image method*, che deduce la direzione e il tempo di arrivo delle prime riflessioni in un punto dell'ambiente, nota la sua topologia;
- le tecniche di *ray tracing*, che modellano le riflessioni successive basandosi sul concetto di *campo diffuso*;
- i metodi statistici, per modellare gli echi che formano il tappeto acustico.

### L'approccio "percettivo"

L'approccio percettivo presenta teoricamente un numero maggiore di desiderabili proprietà rispetto a quello precedente. Il controllo dei parametri percettivi infatti fa sì che l'effetto possa essere adattato al gusto dell'ascoltatore. In più, intuendo che gran parte dell'informazione presente nella risposta impulsiva ambientale non viene utilizzata dai meccanismi uditivi, si può ragionevolmente affermare che un modello percettivo ha maggiori possibilità di essere realizzato efficientemente: al limite, *un* modello in grado di controllare tutti i parametri percettivi può sintetizzare qualunque riverbero.

Le difficoltà sorgono nel momento in cui si cercano di riconoscere dei parametri indipendenti (o *fattori*) capaci di caratterizzare la sensazione del riverbero. In questo senso sono stati fatti molti passi avanti; il paragrafo 6.5 riassume alcune scoperte fatte di recente.

L'obiettivo dei modelli percettivi dunque è l'efficiente realizzazione di un algoritmo di riverbero innanzitutto piacevole all'ascolto, ovvero *naturale*. La strada per raggiungere quest'obiettivo è quella di riprodurre la timbrica del suono lavorando entro la costante di integrazione dell'orecchio, e di generare un valido tappeto acustico.

### Misura dei parametri percettivi

Ricondurre i parametri percettivi a dei valori quantitativi è impresa spesso priva di riscontro valido, se le misure di questi valori non vengono comunque confortate dal parere qualitativo fornito dall'ascoltatore. Ciò nonostante, esistono alcune misure cui fanno riscontro definite impressioni percettive.

Di esse, forse la più significativa è l'*Energy Decay Relief* (EDR) relativo a un ambiente. L'EDR individua una funzione che indicheremo come  $EDR(t, f)$ , la quale offre l'insieme degli andamenti delle curve di decadimento RT al variare della frequenza. Queste curve si possono grossolanamente ottenere sollecitando l'ambiente di volta in volta usando segnali monofrequenziali, e valutando le curve RT ottenute in corrispondenza di ciascuna sollecitazione.

In Figura 6.11 viene riportato l'andamento qualitativo di un diagramma EDR relativo a un'ambiente virtuale. Si noti che  $EDR(0, f)$  fornisce concettualmente l'involuppo della risposta in frequenza dell'ambiente, anche se una misura più significativa di questo involuppo si ha estrapolando funzioni  $EDR(t_0, f)$  per valori fissati di  $t_0$ , in corrispondenza dei quali le prime riflessioni si siano esaurite.

Esperimenti, nei quali si è cercato di mettere in rapporto fattori percettivi indipendenti con grandezze fisicamente misurabili, hanno dimostrato la relazione tra questi fattori e dei corrispondenti rapporti tra pendenze di curve estratte da distinte regioni dell'EDR. Problemi rimasti aperti in questo tipo di esperimenti sono l'individuazione di un numero di fattori percettivi significativi, e la scelta di una risoluzione spatio-temporale per l'EDR in grado di ottimizzare la bontà e l'attendibilità delle relazioni trovate.

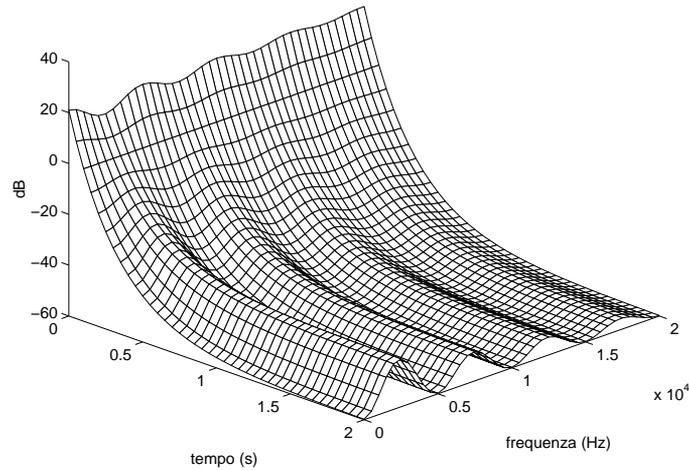


Figura 6.11: Energy Decay Relief: andamento qualitativo.

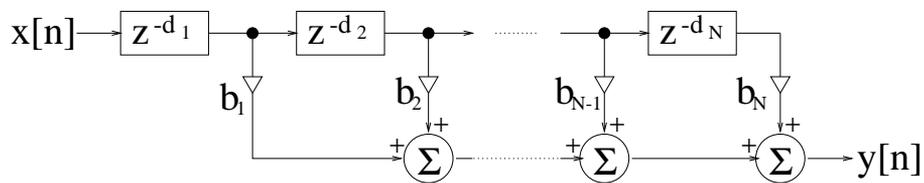


Figura 6.12: Filtro per la simulazione delle prime riflessioni.

### 6.3.4 Realizzazioni

Tutte le considerazioni fatte in precedenza sulla scelta di un modello devono comunque trovare riscontro in strutture realizzabili. La natura del problema conduce essenzialmente a sistemi che fanno uso di filtri lineari nel senso tradizionale del termine, ma non mancano in letteratura realizzazioni più articolate, come le *Feedback Delay Network (FDN)*, cui accenneremo alla fine di questo paragrafo.

#### Riproduzione delle prime riflessioni

Ritornando per un attimo alla Figura 6.9, si può notare che le prime riflessioni possono venire modellate con ottima approssimazione da un filtro FIR. In pratica, la simulazione delle prime riflessioni non scade troppo di qualità se si generano solamente le più significative tra esse, a patto di spaziarle correttamente nell'asse dei tempi.

Un risultato didatticamente interessante è rappresentato nella struttura di Figura 6.12. Essa consiste in un filtro di risposta impulsiva avente trasformata Zeta uguale a

$$H(z) = \sum_{i=1}^N b_i z^{-d_i},$$

e può essere equivalentemente vista come un filtro FIR avente molti coefficienti uguali a 0, tra cui in particolare il primo, o, alternativamente, come una versione generalizzata del FIR, nella quale i ritardi unitari sono sostituiti da linee di ritardo di lunghezza generica  $d_i$ . Non è difficile capire come,

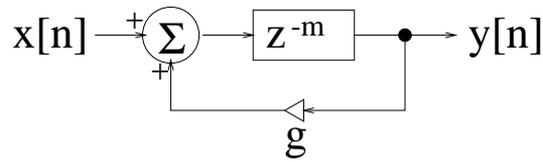


Figura 6.13: Il filtro comb.

immettendo un segnale anecoico in questa struttura, si abbia in uscita un segnale formato dal segnale d'ingresso ritardato più un certo numero di sue ripetizioni pesate, simulanti le prime riflessioni. Il ritardo iniziale, uguale a  $m_1 T$  secondi, modella il tempo di propagazione dalla sorgente al punto di ascolto.

In pratica, può essere necessario ritardare una riflessione di un tempo non multiplo del periodo di campionamento. In tal caso si può ricorrere alle tecniche di interpolazione lineare, viste al paragrafo 6.2.3.

### Riproduzione del campo diffuso

Le formula 6.6 afferma l'esistenza di un'elevata densità degli echi nella regione del campo diffuso. Considerazioni di ordine percettivo confermano l'importanza di conservare questa densità, pena una percezione "granulare" (ovvero *discreta*) di questi echi.

Appare ovvio, in questo caso, l'utilizzo di strutture IIR, capaci di reazionare il segnale in uscita dai filtri. In effetti l'idea più semplice per realizzare economicamente un elevato numero di echi consiste nel processare il segnale mediante un filtro *comb*.

Consideriamo dunque un filtro avente trasformata Zeta uguale a

$$H(z) = \frac{z^{-m}}{1 - gz^{-m}},$$

con  $|g| < 1$ ), come quella di Figura 6.13. La parte iniziale della risposta impulsiva di questo filtro è mostrata in figura 6.14(a), quando è  $g = 1/\sqrt{2}$  e  $M = 5$ . In Figura 6.14(b) compare il modulo della

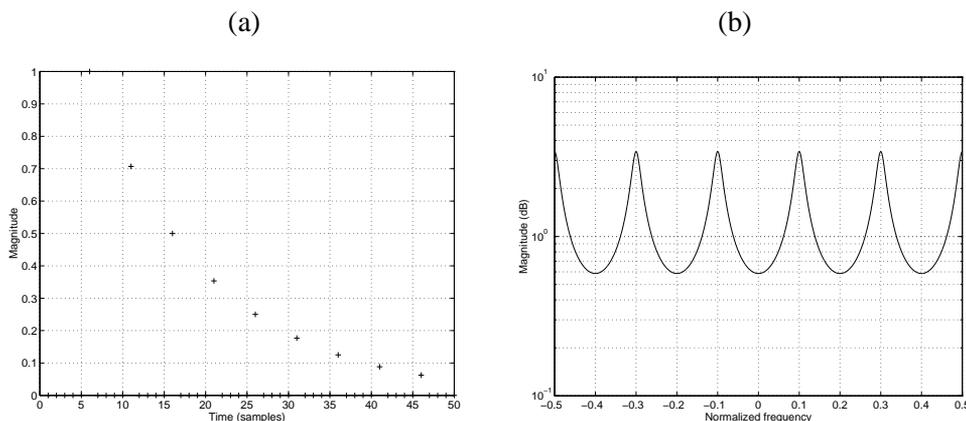


Figura 6.14: (a) risposta impulsiva e (b) risposta in frequenza del filtro comb.

risposta in frequenza, avendo adoperato gli stessi valori per i parametri.

Le figure rendono conto di due proprietà generali del filtro:

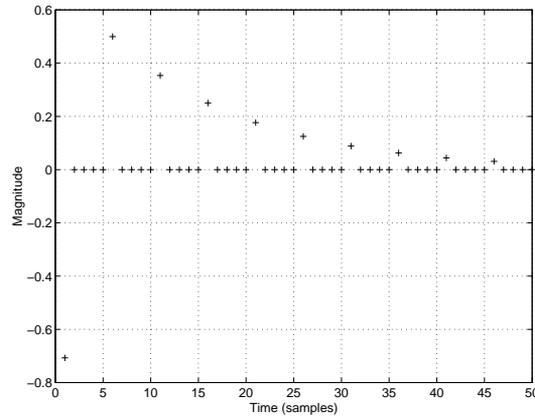


Figura 6.15: Risposta impulsiva del filtro allpass.

1. l' $i$ -esimo eco della risposta impulsiva di un filtro comb, di guadagno di feedback  $g$  ritardo di feedback  $M$ , ha ampiezza uguale a  $g^i$  e occorre in corrispondenza del campione  $im$ ;
2. il modulo della risposta in frequenza dello stesso filtro, valutata tra la continua e la frequenza di Nyquist, presenta  $m$  punti di estremo di quota uguale a  $1/(1-g)$  e  $1/(1+g)$ , alternativamente.

Le posizioni dei punti di estremo superiore, nel modulo della risposta in frequenza, sono individuate dagli argomenti delle posizioni dei *poli* del filtro, ovvero dalle radici  $z_1, \dots, z_m$  del denominatore della trasformata Zeta, risultanti dalla soluzione dell'equazione a valori complessi

$$z = \sqrt[m]{g}.$$

Tenendo conto della condizione imposta dalla (6.7) non è difficile convincersi del fatto che imporre una sufficiente densità degli echi ha, come controparte, l'eccessiva separazione dei picchi in frequenza, nitidamente udibile come una *colorazione* del segnale. Vale ovviamente anche il viceversa. Nè risolve il problema adottare, in luogo del comb, un filtro passatutto (o *allpass*, vedi Capitolo 4, allegato 2, pag. 30) di trasformata Zeta

$$H(z) = \frac{-g + z^{-m}}{1 - gz^{-m}},$$

la cui risposta impulsiva è mostrata in Figura 6.15. Infatti il nostro orecchio è sensibile alle frequenze contenute in brevi finestre di segnale, e da questo punto di vista anche l'allpass, pur contraddistinto dalla piattezza del modulo della risposta in frequenza, colora decisamente il suo ingresso.

Si rende dunque necessario un aumento della complessità della struttura.

Consideriamo un parallelo di  $N$  filtri comb. In questo caso la densità degli echi nel tempo, e parimenti quella dei picchi in frequenza, aumenta linearmente col valore di  $N$ . Infatti la risposta impulsiva del parallelo di filtri comb ha trasformata Zeta uguale a

$$H(z) = \sum_{i=1}^N \frac{z^{-m_i}}{1 - g_i z^{-m_i}},$$

e dunque i poli della struttura complessiva si ricavano dalla

$$H(z) = \prod_{i=1}^N \{g_i - z^{m_i}\} = 0.$$

È sufficiente scegliere i ritardi  $m_1, \dots, m_N$  in modo tale che il massimo comun divisore di questi sia unitario, per essere sicuri di disporre di un sistema la cui risposta in frequenza contiene

$$D_f = m_1 + \dots + m_N$$

picchi nella banda assegnata, e

$$D_t = \sum_{i=1}^N \frac{\prod_{k=1}^N m_k}{m_i} = \sum_{i=1}^N \prod_{k=1, k \neq i}^N m_k$$

echi nel periodo  $T \prod_{i=1}^N m_i$ . Ricordando la (6.8), si può inizialmente dimensionare il sistema, dal punto di vista della densità dei picchi in frequenza, secondo la

$$D_f = \frac{B}{\Delta f_{\max}},$$

dove  $B$  è la banda di utilizzo del filtro. Per quanto riguarda la densità degli echi nel tempo, un dimensionamento di massima va fatto empiricamente.

Affinchè i modi associati a ciascun polo decadano con la stessa legge di decadimento, è necessario che *tutti i poli del sistema siano equidistanti dall'origine degli assi del piano complesso*. Ciò si ottiene semplicemente imponendo la condizione

$$\sqrt[m_i]{|g_i|} = K, \quad i = 1, \dots, N,$$

con  $K$  costante dipendente dal tempo di riverbero ai 60 dB desiderato:

$$K = 10^{-\frac{3T}{RT_{60}}}.$$

Infine, ricordando la Figura 6.15, si noti che applicando  $P$  allpass in cascata al sistema appena visto, si aumenta la densità degli echi di un ulteriore fattore proporzionale a  $P$  senza generare un decadimento delle prestazioni in termini di densità di picchi in frequenza. Dunque, un sistema relativamente semplice ma già in grado di produrre un apprezzabile numero di echi e di picchi può essere quello rappresentato in Figura 6.16.

### 6.3.5 Problemi dei riverberi realizzati con filtri FIR e IIR

Quantunque un riverbero formato dal sistema di Figura 6.12 seguito in cascata da quello di Figura 6.15 presenti già diverse pregevoli caratteristiche, in special modo con riferimento a simulazioni di riverberi di breve durata, diversi artefatti ne limitano la qualità in termini di naturalezza. In particolare:

- la risposta suona inizialmente granulare, in particolare se si devono riverberare suoni brevi e dinamici, ovvero a caratteristica impulsiva;
- successivamente, durante la simulazione delle riflessioni successive, il suono acquista una particolare modulazione (*fluttering*) e, soprattutto, un timbro colorato, metallico.

In effetti, il sistema visto non tiene conto dell'andamento in frequenza del decadimento del suono, visibile anche in Figura 6.11, principalmente dovuto alla caratteristica di assorbimento dell'aria che è più accentuata in alta frequenza. Un semplice modo per affrontare la questione consiste nel sostituire i guadagni di feedback dei filtri comb con dei filtri passabasso. Questa soluzione riduce gli artefatti

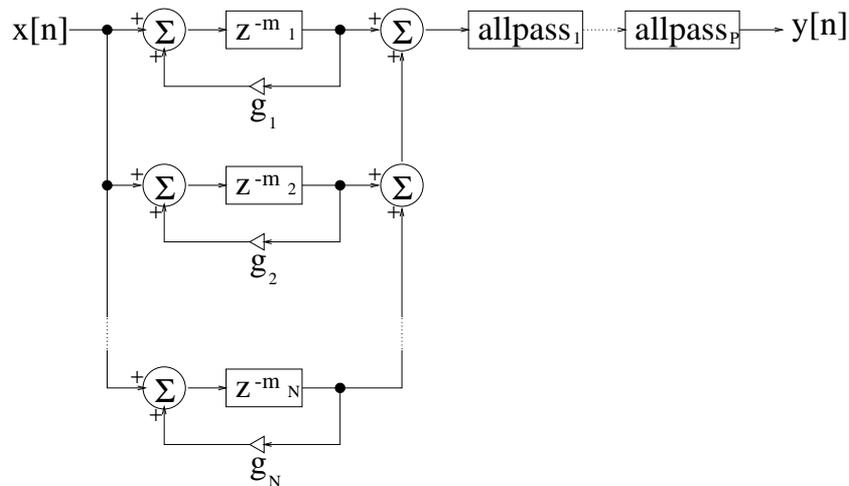


Figura 6.16: Un semplice sistema per la generazione di echi nel tempo e picchi in frequenza.

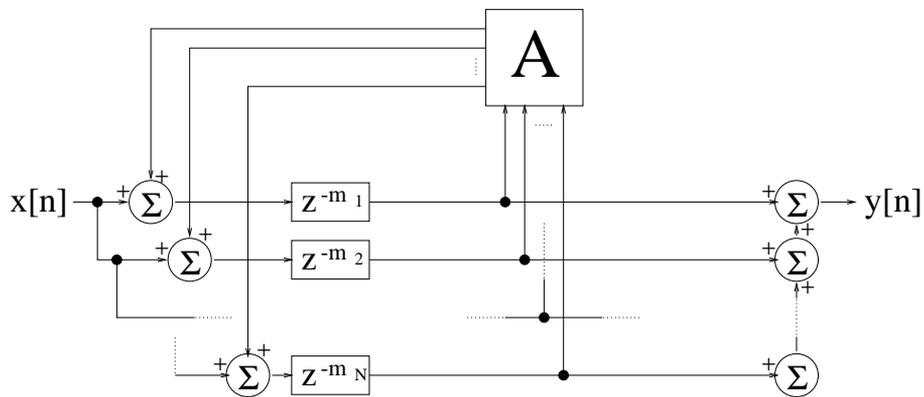


Figura 6.17: Feedback Delay Network.

descritti, ma non li elimina del tutto. In particolare, l'uso di questa tecnica non permette di mettere in relazione le caratteristiche dei passabasso con un corrispondente diagramma EDR.

Altre soluzioni, sicuramente meno giustificabili dal punto di vista teorico ma spesso efficaci, suggeriscono di sostituire le linee di ritardo con linee di ritardo modulate, da applicarsi nei modi e con le cautele viste al paragrafo 6.2.2. In tal caso il tuning dei parametri segue considerazioni prettamente empiriche, legate al gusto dell'ascoltatore e alle capacità del progettista.

Un importante passo avanti è stato compiuto utilizzando le FDN, la cui comprensione offre interessanti spunti di discussione.

### 6.3.6 Feedback Delay Network

Le FDN implementano con la desiderata generalità un sistema LIT. Il segnale viene fatto circolare all'interno di un sistema formato da un filtro multicanale senza memoria e da un parallelo di linee di ritardo, disposti come in Figura 6.17. Il filtro multicanale è completamente descritto da una matrice quadrata  $A$  di dimensione  $N$ , i cui elementi sono semplici coefficienti, detta *matrice di feedback*: indicato con  $s(z)$  il vettore colonna delle trasformate Zeta dei segnali  $s_1, \dots, s_N$  in ingresso al filtro, in uscita dallo stesso si hanno  $N$  nuovi segnali descritti dal vettore  $As(z)$ .

La stabilità è assicurata non appena il determinante di  $A$  è unitario o minore di uno. Una semplice realizzazione prevede la scelta di una matrice contenenti elementi uguali a  $-1$ ,  $0$  e  $1$ , in modo tale che il determinante sia unitario; a questa si antepone un peso  $g$  che osserva la condizione di passività  $|g| < 1$ . In tal modo si assicura peraltro una notevole efficienza in sede di implementazione su di un processore a tempo reale.

Ogni coefficiente di  $A$  non nullo fuori dalla diagonale assicura la commutazione del segnale da un canale d'ingresso al filtro a un diverso canale di uscita dallo stesso. Si noti che, nel caso particolare in cui  $A$  sia uguale alla matrice identità, la FDN si particolarizza in un parallelo di comb come quello di Figura 6.16 (avendo tolto gli allpass), accomunati dall'aver tutti lo stesso coefficiente di feedback,  $g$ .

Le proprietà acustiche di questa realizzazione sono molteplici. In particolare:

- non occorre realizzare la parte del sistema relativa alle prime riflessioni (vedi par. 6.3.4), potendo sostituire l'azione di questo con un'opportuna inizializzazione dello stato della FDN (cioè applicando dei valori iniziali opportuni sulle linee di ritardo). Con questa tecnica risultano peraltro attenuati gli artefatti presenti all'attacco nel sistema visto in precedenza;
- la FDN può prendersi carico di realizzare, oltre al riverbero, delle procedure per la *spazializzazione* del suono.

Col tempo, le FDN sono state migliorate e rese più efficienti, e sono state spunto per il progetto di modelli generalizzati ancora più evoluti. Di questi argomenti diamo cenno nel seguito.

### Evoluzioni delle strutture FDN

Il principale artefatto indesiderato di cui soffre una FDN come quella in Figura 6.17, e che abbiamo già incontrato nel modello di Figura 6.16, è costituito dalla colorazione e dal fluttering presenti nella coda del segnale d'uscita.

Per quanto riguarda il fluttering, una semplice soluzione è quella di sostituire le linee di ritardo della FDN con linee di ritardo modulate tipo quelle del paragrafo 6.2.2 le quali, come sappiamo, richiedono una taratura empirica.

Riguardo la colorazione anche qui, come nel caso dei filtri comb, si può pensare di applicare dei filtri passabasso, immediatamente a valle delle linee di ritardo. In queste strutture tuttavia viene elegantemente risolto il problema di *progettare le caratteristiche di assorbimento dei filtri passabasso sulla base di un EDR desiderato*. La tecnica, di cui diamo qui un brevissimo cenno, consiste nel progettare un prototipo FDN senza perdite (*lossless prototype*), ovvero una FDN con retroazione a guadagno unitario e la cui risposta all'impulso sia essenzialmente un segnale privo di colorazione — per inciso, una FDN con queste caratteristiche si può progettare adoperando matrici di feedback molto efficienti dal punto di vista del carico computazionale — e, successivamente, nell'applicare a valle delle linee di ritardo dei filtri specificati in base a una funzione  $EDR(t, f)$  desiderata. Detti filtri peraltro risultano semplici ed efficienti. Con questa tecnica è possibile raggiungere dei risultati molto buoni in termini di naturalezza del riverbero.

I modelli FDN sono stati ulteriormente generalizzati, sostituendo il classico concetto di segnale con quello di *segnale d'onda* (*wave signal*): alle linee di trasmissione del segnale (monodirezionali) vengono sostituite delle guide d'onda a tempo discreto (*digital waveguides*, o DW), nelle quali il segnale  $s$  viene modellato, in base alla fisica delle onde, come la sovrapposizione di due segnali d'onda  $s_+$  e  $s_-$  che si propagano in opposte direzioni: le DW, dunque, trasmettono informazioni bidirezionalmente.

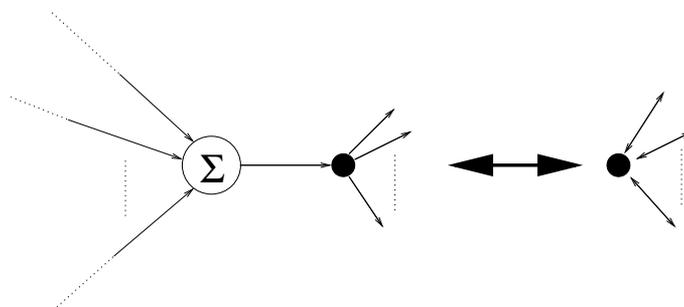


Figura 6.18: Propagazione di un segnale secondo l'approccio tradizionale, a sinistra, e interpretazione della propagazione secondo il modello *waveguide*, facendo uso di DW e di giunzioni di scattering, a destra.

Questa posizione permette di generalizzare la matrice di feedback in una *matrice di scattering* e, più in generale, permette di modellare tutti i punti dove un segnale viene suddiviso in più canali o, viceversa, dove più segnali vengono sommati a formare un'uscita, con delle *giunzioni di scattering* (vedi Figura 6.18).

La struttura risultante dalla generalizzazione della FDN viene chiamata *Digital Waveguide Network* (DWN), e racchiude una serie di proprietà più generali di quelle proprie della FDN.

## 6.4 Spazializzazione

L'ascolto binaurale di una sorgente monofonica fornisce all'ascoltatore, oltre all'informazione relativa al messaggio sonoro emesso dalla sorgente, una serie di percezioni legate alle caratteristiche acustiche dell'ambiente d'ascolto. L'ascoltatore, in altre parole, identificherà non solo il suono originale ma anche, possibilmente, la direzione di provenienza del messaggio sonoro, le dimensioni della sorgente, le caratteristiche dell'ambiente d'ascolto.

La complessità delle informazioni di cui si arricchisce un suono puro, durante il suo tragitto dalla sorgente alle orecchie dell'ascoltatore, determina la *spazializzazione* che viene apportata al segnale. In analogia a quanto visto in precedenza, scopo di questa sezione è la ricerca di validi modelli per la spazializzazione.

Appare evidente da subito che l'argomento spazializzazione include il riverbero, già trattato alla sezione 6.3. Qui ci si concentrerà sui metodi per creare, da un suono puro, una coppia di segnali che, ascoltati senza ulteriori modifiche, siano in grado di suscitare una convincente impressione di localizzazione e di "presenza" di una sorgente sonora virtuale.

I sistemi in grado di assolvere a questo compito prendono il nome di *binauralizzatori*, e dunque a rigore sarebbe stato più opportuno intitolare la presente sezione col nome di binauralizzazione. Chiarito una volta per tutte che riverbero e binauralizzazione sono entrambi compendati nella spazializzazione, la terminologia pratica predilige la scelta qui fatta, e comunque una breve trattazione di psicoacustica della spazializzazione viene fatta alla sezione 6.5.

Nel proseguo, vedremo dapprima alcuni concetti e strumenti preliminari: il modello HRTF, la KEMAR, le metodologie di approccio ai modelli. Successivamente si presenterà un modello strutturale per la binauralizzazione.

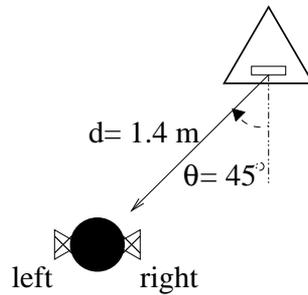


Figura 6.19: Misura di HRTF. Le risposte  $h_l$  e  $h_r$  vengono valutate in prossimità dei punti indicati con  $\times$ .

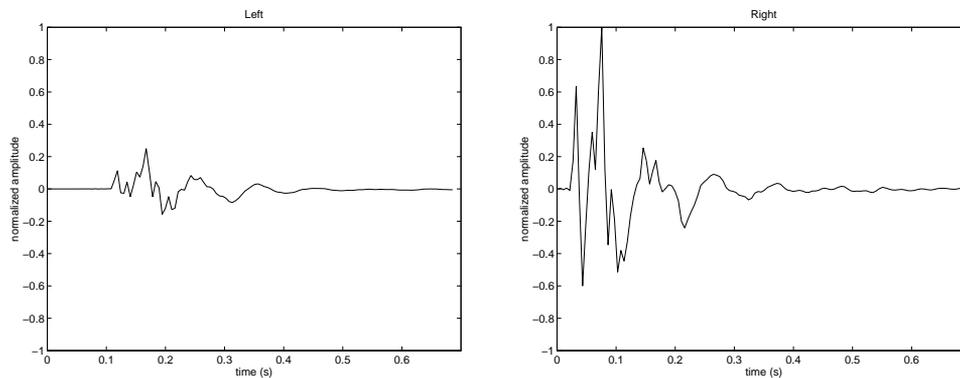


Figura 6.20: Andamenti nel tempo delle HRIR misurate nella condizione sperimentale di Figura 6.19.

### 6.4.1 Il modello HRTF

Si supponga di porsi in un ambiente anecoico, nei pressi di una sorgente sonora posta a una certa distanza dalla testa. L'ascolto *binaurale* della sorgente monofonica è responsabile delle informazioni spaziali (non legate al riverbero) presenti nel messaggio sonoro udito. Il modello HRTF attribuisce queste informazioni alle due risposte impulsive  $h_l$  e  $h_r$ , denominate HRIR (*Head Related Impulse Responses*), misurate dal punto di emissione del suono verso i punti di ricezione posti all'interno dei padiglioni auricolari, rispettivamente sinistro e destro. Ovviamente la stessa informazione è presente nelle Funzioni di Trasferimento dedotte dalle HRIR rispettivamente, e denominate appunto HRTF (*Head Related Transfer Functions*).

La misura delle HRTF viene schematizzata nella Figura 6.19, dove è presentata una condizione sperimentale nella quale la sorgente è distante  $d = 1.4$  m dalla testa, spostata rispetto a questa sul piano orizzontale di un *azimuth*  $\theta = 45^\circ$ , e sul piano verticale di un'elevazione  $\phi = 0^\circ$ .

In Figura 6.20 si presentano gli andamenti delle HRIR misurate nelle condizioni sperimentali relative all'esempio sopra. Alcune interessanti considerazioni, peraltro necessarie per motivare il prossimo paragrafo, sono d'obbligo:

- l'attacco del segnale viene mascherato quasi del tutto all'orecchio sinistro. Quest'effetto è noto come l'"ombra della testa" (*head shadow*), il quale fa da filtro soprattutto per le alte frequenze, di cui appunto è ricco l'attacco;
- riflessioni residue del segnale, causate principalmente dal torso e dalle spalle, giungono a entrambe le orecchie;

- la componente dominante di bassa frequenza, ben visibile tra 0.25 s e 0.4 s, giunge pressochè identica a entrambe le orecchie, a meno di uno sfasamento dipendente dalla distanza interaurale. Questo poichè alle basse frequenze i fenomeni di diffrazione dell'onda prevalgono su quelli di mascheramento.

Test nei quali sono stati fatti ascoltare in cuffia segnali anecoici o pre-riverberati, filtrati adoperando opportune HRTF, attestano la bontà del modello. In perfetta analogia con quanto detto nel caso del riverbero (vedi paragrafo 6.3.3), la realizzazione di un binauralizzatore mediante la costituzione di un database di HRTF, invocate magari sulla base della posizione — rilevata mediante sensori — della testa, solleva problemi di costo del modello. A favore di questo approccio giocano comunque studi recenti che hanno dimostrato la bassa sensibilità del nostro apparato uditivo al dettaglio spettrale delle HRTF: nello specifico è stato osservato come una semplificazione degli spettri, ottenuta imitandone l'andamento solamente macroscopico del modulo e senza imporre alcun vincolo alla fase, produca al più una variazione dell'elevazione della sorgente acustica virtuale.

#### 6.4.2 IID e ITD

Le leggi fisiche dell'interferenza sulla propagazione delle onde acustiche causata da un ostacolo, riassunte per i nostri scopi dalle considerazioni fatte commentando le risposte impulsive di Figura 6.20, unite a evidenze emerse durante esperimenti psicoacustici sulla percezione spaziale, autorizzano a postulare l'esistenza, a livello neurologico, di due apparati deputati rispettivamente alla misura della *differenza d'intensità interaurale* (*interaural intensity difference*, IID), nota anche come ILD (*interaural level difference*), e della *differenza di tempo interaurale* (*interaural time difference*, ITD).

- Il primo apparato registra la differenza delle energie nei segnali che giungono alle orecchie. È specialmente utile quando il messaggio acustico contiene alte frequenze (tipicamente al di sopra di 1500 Hz) le quali, come abbiamo visto, risentono del mascheramento causato dall'effetto di *head shadow*.
- Il secondo apparato valuta le differenze di fase tra i segnali. Poichè tali differenze perdono di significato al di sopra di una certa frequenza, la misura dell'ITD viene eseguita specialmente sulle componenti di bassa frequenza.

Il meccanismo di misura dell'IID e dell'ITD pare essere alla base delle nostre facoltà di discernere la direzione di provenienza di un messaggio sonoro, assieme ad altri importanti parametri illustrati più diffusamente alla sezione 6.5.

#### 6.4.3 La testa KEMAR

Il modo più comodo per misurare le HRIR è quello di adoperare una testa sintetica (*dummy head*), di solito la KEMAR. In effetti, il padiglione auricolare (o *pinna*) è altamente responsabile della nostra capacità percettiva spaziale, e dunque il suo contributo va integrato nell'informazione propria delle HRTF. Di fatto, ogni modello per la binauralizzazione, in grado di conservare la morfologia o quantomeno alcune proprietà delle HRTF, non può prescindere dall'esigenza di modellare la pinna.

Ciò implica che eventuali microfoni per la misurazione delle HRTF vanno collocati all'interno del condotto uditivo: di qui si comprende l'utilità di disporre di una testa sintetica.

Ancora, la testa KEMAR si rende utile quando si vogliono determinare le HRTF private del contributo informativo aggiunto dai padiglioni auricolari che, essendo nel caso della KEMAR rimovibili, permettono di eseguire queste particolari misure.

#### 6.4.4 Modelli

Come già detto, un approccio squisitamente fisico al problema conduce a modelli molto dispendiosi in termini di memoria e potenza di calcolo richieste. Purtroppo, a differenza del caso del riverbero, gli studi condotti nel campo della spazializzazione non sono riusciti a determinare dei parametri percettivi al punto di poter sviluppare, mediante questi, dei modelli efficaci, anche se la natura del problema suggerisce che in realtà pochi parametri (raggio ed eccentricità della testa, diametro della pinna ecc...) determinano il risultato. In tal senso, una complicazione è causata dall'elevato grado di *soggettività* delle sensazioni spaziali: un sistema ben tarato sui parametri di un singolo ascoltatore in generale non si può esportare a un altro ascoltatore.

Disattesa per il momento la speranza di estrarre dei parametri percettivi dall'informazione contenuta nelle HRTF, la scelta del modello deve dunque restare sostanzialmente all'interno dell'approccio fisico. Alcune soluzioni proposte sono elencate di seguito:

- *modelli zero-polo*: si cerca con essi di simulare le HRTF sostanzialmente con delle FdT semplificate. Di fatto, funzioni capaci di produrre dei convincenti effetti spaziali sono in genere caratterizzate da un numero di coefficienti che non porta a realizzazioni particolarmente efficienti;
- *espansioni in serie*: le HRTF possono venire matematicamente rappresentate adoperando basi alternative di funzioni ortogonali, pesate da coefficienti dipendenti dalla posizione relativa sorgente-punto d'ascolto. Si ottengono in questo modo modelli più efficienti, non abbastanza tuttavia per realizzare in tempo reale spazi virtuali variabili con la posizione angolare della testa;
- *modelli strutturali*, i quali cercano di simulare pari pari i principali elementi che il segnale acustico incontra durante il suo cammino dall'ambiente fino al condotto auricolare: spalle, testa e pinna. Sono questi i modelli che hanno prodotto i risultati più interessanti, permettendo buone riproduzioni della sensazione spaziale pur rimanendo entro un contenuto costo computazionale.

Nel seguito si presenta uno di questi modelli strutturali, interessante sia per chiarezza didattica che per efficacia.

#### 6.4.5 Un modello strutturale

Un modello strutturale cerca di simulare il contributo informativo introdotto nel segnale da parte di una struttura, che nel nostro caso è composta dalla seguente serie di blocchi (Figura 6.21):

1. spalle (che causano la formazione di echi);
2. testa (che contribuisce con l'*head shadow*);
3. pinna (che causa ulteriori echi).

Il nostro modello dedurrà i singoli blocchi, ne valuterà l'importanza in termini di contributo informativo, infine definirà un sistema capace di implementare efficientemente il modello. Allo scopo sarà utile disporre di una simbologia che definisca i segnali in base ai blocchi — S (spalle), T (testa) e P (pinna) — attraversati: detto  $x$  il segnale in ingresso, chiameremo ad esempio  $x_{ST}$  il segnale che ha attraversato il blocco relativo alle spalle e quello relativo alla testa, e  $x_{SP}$  un segnale processato dal blocco delle spalle e da quello relativo alla pinna, privo del contributo informativo dato dall'*head shadow*.

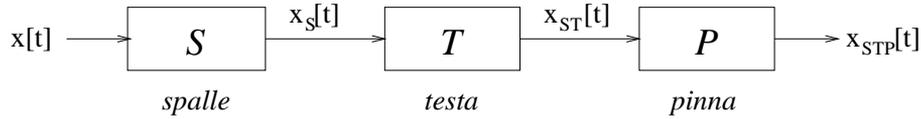


Figura 6.21: Un modello di struttura per la spazializzazione.

### Deconvoluzione della head shadow

La possibilità di modellare il blocco T, ovvero la *head shadow*, mediante un filtro a tempo continuo  $H_T$  a fase minima, caratterizzato cioè dall'aver una FdT i cui zeri siano tutti a parte reale negativa, permette di deconvolvere dalle HRTF il contributo — modellato dal filtro — della testa. Ciò si ottiene filtrando le HRTF con l'inverso del filtro *head shadow*,  $1/H_T$ , che esiste ed è stabile per la proprietà di minimalità della fase. Il contributo di questo filtraggio dunque va a elidersi con il contributo del blocco T.

Il modello che analizziamo suppone l'esistenza di tale filtro. Risolvendo il problema della diffrazione di un'onda su un corpo rigido sferico, si può dedurre un semplice filtro a fase minima del primo ordine che, completato ponendo in cascata a esso un ritardo, simula egregiamente la soluzione esatta al problema.

Fatti i dovuti calcoli, la FdT del filtro analogico risulta essere uguale a

$$H_T(\omega, \theta) = \frac{1 + j \frac{\alpha(\theta)\omega}{2\omega_0}}{1 + j \frac{\omega}{2\omega_0}} e^{-j\omega\Delta T(\theta)}, \quad (6.9)$$

con

$$\alpha(\theta) = \left(1 + \frac{\alpha_{\min}}{2}\right) + \left(1 - \frac{\alpha_{\min}}{2}\right) \cos\left(\frac{\theta}{\theta_{\min}}\pi\right) \quad (6.10)$$

e con

$$\Delta T(\theta) = \begin{cases} -\frac{1}{\omega_0} \cos\left(\frac{\theta}{180^\circ}\pi\right) & , \quad 0 < |\theta| < 90^\circ \\ \frac{1}{\omega_0} \frac{|\theta| - 90^\circ}{180^\circ} \pi & , \quad 90^\circ \leq |\theta| < 180^\circ \end{cases} \quad (6.11)$$

Posto

$$\begin{aligned} \alpha_{\min} &= 0.1 \\ \theta_{\min} &= 150^\circ \\ \omega_0 &= c/a, \end{aligned} \quad (6.12)$$

con  $a$  raggio della sfera e  $c$  velocità del suono in aria, e rispettate dunque le condizioni  $\omega_0 > 0$  di stabilità e  $\alpha > 0$  di minimalità della fase, si ottiene un filtro — dipendente dall'azimuth  $\theta$  relativo alla posizione d'ascolto rispetto alla sorgente — che approssima in maniera soddisfacente la FdT esatta.

La dimostrazione della stabilità e della minimalità della fase del filtro  $H_T$ , privato del contributo dato dal ritardo (6.11), è riportata nel paragrafo 6.4.5.

Il modello appena ottenuto viene verificato in un primo momento deconvolvendo le risposte misurate ai due condotti uditivi della testa KEMAR senza le pinne. La somiglianza delle due funzioni, formalmente corrispondenti ai segnali  $x_S$  relativi al lato sinistro e destro, sancisce l'attendibilità dell'informazione fornita da queste sul contributo del busto e soprattutto delle spalle. La loro semplicità autorizza a valutarle osservandone solo l'andamento nel tempo: i grafici, che qui non vengono riportati, evidenziano abbastanza chiaramente l'eco dovuto al contributo delle spalle, in buona sostanza unico, che arriva successivamente al primo picco causato dall'arrivo del segnale diretto.

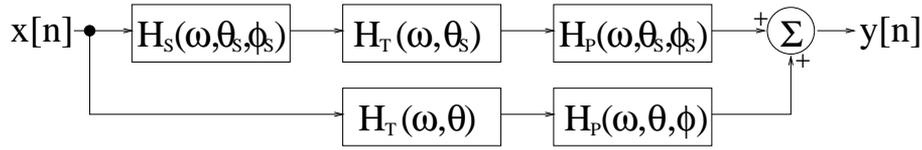


Figura 6.22: Modello strutturale per la percezione di eventi spaziali.

### Contributo della pinna

Reinserendo le pinne, si osserva il loro contributo sotto forma di echi supplementari presenti nei segnali  $x_{SP}$ . Le pinne hanno interessanti conseguenze sulla percezione dell'elevazione della sorgente acustica. Infatti ruotando la KEMAR in senso verticale si registra uno spostamento degli echi dovuti alle pinne; viceversa, il loro contributo all'informazione relativa all'azimuth non appare determinante.

Dall'osservazione diretta dei due segnali  $x_{SP}$  si possono apprezzare, mediante confronto con i rispettivi segnali  $x_S$ , gli echi introdotti dalle pinne. Questi possono essere direttamente modellati attraverso filtri FIR, che tengano almeno conto dei valori di picco dei principali echi, e delle posizioni relative di questi. Il risultato sarà un filtro avente la seguente FdT:

$$H_P(\omega, \theta, \phi) = 1 + \sum_{n=1}^L \rho_i(\theta, \phi) e^{j\omega\tau_i(\theta, \phi)}, \quad (6.13)$$

dependente dall'azimuth e dall'angolo di elevazione  $\phi$  della sorgente rispetto alla testa. Si ricorrerà all'interpolazione lineare (vedere paragrafo 6.2.3) quando uno o più ritardi  $\tau_i$  non siano multipli del periodo di campionamento del sistema.

### Struttura completa del modello

Accorpendo tutte le osservazioni fatte, si può strutturare il modello come appare in Figura 6.22. Il suono diretto viene in primo luogo riflesso dalla spalla; questo fenomeno viene modellato ponendo in parallelo al segnale diretto un singolo eco, descritto dalla FdT

$$H_S(\omega, \theta_S, \phi_S) = \rho_S(\theta_S, \phi_S) e^{j\omega\tau_S(\theta_S, \phi_S)},$$

nella quale il pedice S nei parametri di azimuth ed elevazione tiene conto del diverso angolo di ingresso al padiglione auricolare. Successivamente, i segnali incontrano la *head shadow*, formalizzata dalla (6.9) che varia in base all'azimuth. Infine, la pinna contribuisce ad un ulteriore processamento del segnale, mediante il modello espresso dalla (6.13).

Si rammenta che la struttura vista spazializza il segnale solo se viene applicata, con le opportune variazioni parametriche, a entrambe le orecchie. Dunque, il binauralizzatore complessivo sarà costituito dal parallelo formato da due strutture come quelle di Figura 6.22, aventi un ingresso monofonico comune.

### Implementazione del modello

Ulteriori esperimenti hanno evidenziato la non assoluta necessità di modellare la riflessione dovuta alla spalla. La possibilità di tagliare il canale ove è presente la FdT  $H_S$  si traduce in una semplificazione del modello.

$n$	$\rho_n$	$A_n$	$B_n$	$D_n$
1	0.5	1	2	0.85
2	-1	5	4	0.35
3	0.5	5	7	0.35
4	-0.25	5	11	0.35
5	0.25	5	13	0.35

Tabella 6.1: Valori numerici indicativi dei parametri del modello della pinna.

Simulazioni soddisfacenti sono state ottenute adoperando, per il calcolo del tempo d'arrivo dell' $n$ -esimo eco dovuto alla pinna, la seguente funzione:

$$\tau_n(\theta, \phi) = A_n \cos\left(\frac{\theta}{360^\circ}\pi\right) \sin\left[D_n \frac{90^\circ - \phi}{180^\circ}\pi\right] + B_n, \quad n = 1, \dots, L,$$

peraltro periodica nelle variabili di azimuth ed elevazione, come sarebbe da aspettarsi. Essa dà buoni risultati per  $-90^\circ \leq \theta \leq 90^\circ$  e  $-90^\circ \leq \phi \leq 90^\circ$ .

Si è anche verificata la sostanziale indipendenza dall'azimuth e dall'elevazione dei coefficienti di riflessione  $\rho_i$ . Infine, si è deciso di assegnare una volta per tutte il numero di echi causati dalla pinna:  $L = 5$ .

Fatte queste semplificazioni, la scelta definitiva dei parametri  $\rho_n$ ,  $A_n$ ,  $B_n$  e  $D_n$  va fatta soggettivamente. Valori di riferimento, risultati efficaci per alcuni soggetti, sono illustrati in Tabella 6.1.

### Risultati e punti di discussione

Il modello strutturale visto è stato testato in condizioni tali da concentrare il più possibile l'attenzione sulle sue proprietà di spazializzazione. Adoperando suoni atimbrici (rumore gaussiano nel caso particolare) si è verificato che la capacità di riconoscere spazi acustici virtuali, dall'ascolto di suoni filtrati da HRTF "personalizzate", statisticamente non si discosta di molto dalla capacità di riconoscere gli stessi spazi quando alle HRTF venga sostituito il modello strutturale, anche questo tarato soggettivamente.

Restano aperti alcuni punti:

- il modello pare evocare esternalizzazione. Ciò non deve sorprendere quando si pensi che le HRTF non portano alcuna informazione — ad esempio sotto forma di riverbero — circa le caratteristiche dell'ambiente (si veda in particolare la sezione 6.5), e questo comporta la difficoltà non già di percepire la direzione di provenienza di un suono, bensì di collocare la sorgente sonora all'interno di un ambiente;
- gli artefatti sul timbro che il modello eventualmente apporta sono da valutare nelle diverse condizioni d'ascolto;
- FdT "residue", come quelle esistenti tra l'altoparlante di una cuffia e l'ingresso all'orecchio, possono in linea di principio introdurre a loro volta degli artefatti dei quali occorre eventualmente tenere conto. Una possibile soluzione a questo punto consiste nell'adottare speciali auricolari che evitano qualunque interferenza tra altoparlanti e orecchio esterno. È chiaro però che una soluzione come questa non appare praticabile specialmente quando il modello persegue degli scopi commerciali.

### Deduzione della stabilità e della minimalità della fase del filtro head shadow

Si consideri la FdT (6.9), privata della componente data dal ritardo (6.11) e nella quale siano stati sostituiti i parametri dati dalle (6.10) e (6.12).

Conduciamo dapprima una verifica di stabilità della (6.9), ovvero stabiliamo se il suo prolungamento analitico al piano complesso definisce una funzione, nella variabile complessa  $s$ , i cui poli giacciono nel semipiano sinistro. È:

$$\frac{1 + j \frac{\alpha(\theta)\omega}{2\omega_0}}{1 + j \frac{\omega}{2\omega_0}} \xrightarrow{(j\omega \rightarrow s)} \frac{1 + \frac{\alpha(\theta)s}{2\omega_0}}{1 + \frac{1}{2\omega_0}s},$$

la quale possiede evidentemente un polo posto in  $s = -2\omega_0$ . Per la (6.12), la stabilità della (6.9) è assicurata.

Per quanto riguarda la minimalità della fase, essa sussiste se e solo se lo zero della funzione prolungata per analiticità giace sul semipiano sinistro del piano complesso. Non è difficile osservare che lo zero verifica la condizione

$$s = -\frac{2\omega_0}{\alpha(\theta)}$$

da cui, assunta la stabilità, discende la fase minima non appena sia verificata la condizione  $\alpha > 0$ . Ciò è vero quando ci si attiene alla scelta fatta in (6.10).

## 6.5 Appendice: percezione di eventi spaziali

Ancorchè non indispensabile per la comprensione dei sistemi presentati alle Sezioni precedenti, questa appendice completa tutta la trattazione svolta, cercando di chiarire, pur molto sinteticamente, alcuni aspetti psicoacustici legati alla percezione spaziale. Si è preferito inserire questi concetti dopo i fatti più tecnici in quanto essi, pur non fornendo direttamente dei parametri di progetto, devono tuttavia completare la base di conoscenze proprie del progettista di questo genere di sistemi, siano essi un chorus, un riverbero o un binauralizzatore.

### 6.5.1 Terminologia

Nel tentativo di definire le sensazioni spaziali legate alla percezione dei suoni, la vasta letteratura esistente in materia presenta diverse terminologie, a cui fanno riscontro almeno un pari numero di tipi di sensazioni avvertibili. Non di rado, i concetti alla base delle terminologie coniate sono entrati in parziale sovrapposizione tra loro, rendendo ardua una definitiva classificazione delle sensazioni spaziali. Del resto, la psicoacustica della percezione spaziale non ha fortunatamente ancora avuto la presunzione di proporsi come una scienza esatta.

Ad ogni modo, la ricerca pare orientarsi progressivamente verso l'analisi di quattro aspetti percettivi:

1. l'*impressione spaziale* (*spatial impression*, SI), ovvero la capacità di evincere un'informazione spaziale da quella sonora (ad esempio, percepire di trovarsi in una sala da concerto o in una stanza d'ufficio);
2. la *spaziosità* (*spaciousness*), conosciuta anche come *inviluppo* (*envelopment*), ovvero la percezione dell'ampiezza dell'ambiente acustico (ad esempio, percepire di trovarsi in una sala da concerto e non in una stanza d'ufficio);

3. la *dimensione apparente della sorgente* (*apparent source width*, ASW), ovvero la percezione delle dimensioni della sorgente sonora (ad esempio la capacità di distinguere un diffusore da una grande orchestra);
4. la posizione relativa della sorgente sonora, ovvero la capacità, propria dell'ascoltatore, di collocare la posizione della sorgente sonora all'interno di uno scenario tridimensionale. Ciò equivale a individuare un angolo orizzontale (o *azimuth*) e un angolo verticale (o *elevazione*) in relazione alla direzione di provenienza del suono.

Diversi esperimenti sono stati condotti nell'intento di motivare questi aspetti sulla base di considerazioni legate ai parametri architettonici e acustici dell'ambiente d'ascolto, e alle caratteristiche temporali e spettrali della coppia di segnali che giungono alle orecchie; dunque, considerazioni di tipo fisico. Se da una parte questi esperimenti hanno spesso successo nelle perlopiù semplificate condizioni sperimentali allestite durante un test, d'altra parte le conclusioni che ne seguono entrano non di rado in contraddizione con i risultati di altri esperimenti, o con evidenze riscontrabili in particolari situazioni d'ascolto.

Preso atto dei limiti riscontrati in un approccio di tipo fisico al problema, la ricerca non ha tuttavia trascurato l'approccio percettivo. Nonostante le problematiche legate allo scollamento di quest'approccio da alcuni aspetti tecnico-progettuali, in analogia con quanto illustrato al paragrafo 6.3.3, esso conduce in generale a modelli in grado di interpretare meglio la percezione spaziale del suono.

### 6.5.2 Ipotesi per un modello percettivo

Uno schema accettato dalla maggior parte degli studiosi prevede che l'informazione sonora, suddivisa in bande critiche (vedi paragrafo 2.9), giunga in particolare alle sedi neurologiche preposte alla valutazione dell'IID e dell'ITD, già definiti al paragrafo 6.4.2. Confrontando diversi studi, pare probabile che queste sedi siano concettualmente distinte, cioè che le informazioni estratte dall'IID e dall'ITD, non necessariamente in accordo tra loro, vengano separatamente valutate.

A questo punto, l'informazione appena acquisita (sempre suddivisa in bande) va a quantificare tre parametri:

1. un parametro di *direzionalità*, legato alla provenienza dell'informazione;
2. un parametro di *intensità*, legato alle energie del messaggio informativo;
3. un parametro di *fluttuazione spaziale*, legato alle variazioni di energia e punto di provenienza del messaggio informativo. Questo parametro discende da considerazioni legate alle caratteristiche dei segnali che giungono alle orecchie: si dimostra infatti che la sovrapposizione di un segnale monofrequenziale diretto e di una sua riflessione vanno a formare un nuovo segnale la cui fluttuazione spaziale angolare media attorno a un punto di provenienza (chiamata anche fluttuazione dello *pseudoangolo*) varia con l'energia e con l'angolo d'incidenza del segnale riflesso rispetto al segnale diretto. Variando la frequenza, cambia la legge che lega la fluttuazione dello pseudoangolo all'energia e all'angolo d'incidenza.

Questi parametri vengono utilizzati per riconoscere nel segnale acustico degli "eventi" (*phones*), che vengono associati alternativamente a un "flusso in primo piano" (*foreground stream*), corrispondente all'insieme dei messaggi acustici che attribuiamo essere provenienti direttamente dalla sorgente sonora, e a un "flusso di *background*" (*background stream*), corrispondente a quello che il nostro senso dell'udito generalmente identifica come riverbero.



Eventi che vengono attribuiti al flusso in primo piano determinano nei primi istanti di ascolto (fino a circa 50 ms) la localizzazione della direzione di provenienza del suono e l'ASW: tanto più precisa la prima quanto minore la seconda e viceversa, ciò dipendendo dalla precisione di identificazione dell'informazione da parte dei meccanismi di IID e di ITD, dunque dalla coerenza dell'informazione fornita dai due meccanismi nelle diverse bande critiche. Successivamente (dopo circa 50 ms) la fluttuazione dello pseudoangolo fornisce l'informazione per determinare l'impressione spaziale.

Il CSI viene evocato da segnali percettivamente continui. Brevi segnali impulsivi creano stimoli coerenti e di semplice decodifica per gli apparati IID e ITD, utili soprattutto per l'identificazione della direzione di provenienza del suono. Passando dal carattere impulsivo a quello continuo, l'aumento progressivo della fluttuazione dello pseudoangolo, dovuta alla presenza delle riflessioni del segnale continuo, abbassa la precisione nell'identificazione della direzione di provenienza e, per quanto detto, aumenta l'ASW e definisce l'impressione spaziale. Si noti tuttavia come l'aumento della fluttuazione non avvenga necessariamente, ciò influenzando sul tipo di sensazione evocata. Si noti anche la possibilità per il CSI di assumere caratteristiche "inviluppanti", non appena inizi ad assomigliare a un segnale riverberante.

L'ESI viene evocata in un contesto differente da quello ove prevale il CSI, ovvero in occasione di eventi di breve durata (dell'ordine di 50 ms) costituiti da un segnale e dalle sue prime riflessioni, di ampiezza confrontabile con quella del segnale diretto. I fenomeni che avvengono in una finestra temporale così breve vengono raggruppati dall'orecchio in un unico evento. Questo evento fornisce tipicamente sensazioni di ASW e di SI, ma non di inviluppo, ed è tipico dei piccoli ambienti.

Le fluttuazioni dello pseudoangolo legate a questi eventi producono effetti non facilmente predicibili. Solitamente un aumento della fluttuazione produce un allargamento dell'ASW, e modificazioni di timbro, mentre la posizione relativa della sorgente resta di solito legata alla direzione di provenienza del segnale diretto. Piccoli ambienti dotati di pareti altamente riflettenti costituiscono interessanti casi particolari, in grado al limite di evocare anche sensazioni di inviluppo.