

Capitolo 5

Sintesi dei segnali audio

Giovanni De Poli

Carlo Drioli

Federico Avanzini

Copyright ©1999 by Giovanni De Poli, Carlo Drioli and Federico Avanzini.
All rights reserved.

5.1 Introduzione

Negli strumenti musicali tradizionali il suono è prodotto dalla vibrazione di parti meccaniche. Negli strumenti sintetici, la vibrazione è descritta da funzioni nel tempo, dette segnali, che esprimono la variazione nel tempo della pressione acustica.

Per costruire uno strumento musicale tradizionale è sufficiente realizzare un dispositivo che sfrutta uno dei tanti meccanismi fisici per la produzione di vibrazioni. Negli strumenti musicali sintetici, invece, l'obiettivo è di generare una più astratta funzione nel tempo, detta segnale acustico. Allo scopo è necessario implementare una rappresentazione semplificata ed astratta del modo di produzione del suono, detta *modello*. Il modello del segnale, quindi, gioca il ruolo del fenomeno fisico negli strumenti tradizionali: esso costituisce il nucleo centrale attorno al quale si sviluppa la realizzazione dello strumento musicale. Nel modello l'astrazione ha il significato di inscrivere il meccanismo di produzione in una classe più generale di cui esso rappresenta un caso particolare. La semplificazione tende a focalizzare la descrizione sugli aspetti ritenuti significativi nel caso in esame. Spesso nella descrizione del modello si ricorre a relazioni matematiche per legare le cause con gli effetti; in questo modo, usando il modello si può prevedere il comportamento del fenomeno in condizioni note. Le condizioni note includono i parametri del modello, eventualmente variabili nel tempo, e lo stato iniziale da cui procede l'evoluzione.

L'*algoritmo* è il procedimento effettivo che consente di realizzare ciò. L'implementazione dell'algoritmo su un processore consente di calcolare l'evoluzione del fenomeno, eventualmente nella sua forma semplificata. In particolare algoritmi di sintesi dei segnali musicali, basati su modelli del segnale che si focalizzano su diverse e interessanti proprietà degli stessi, consentono di ottenere l'evoluzione temporale del segnale. Al variare dei parametri si ottengono tutti i possibili segnali della classe identificata dal modello; è cioè possibile esplorare l'insieme dei timbri descritto dal modello.

In linea di principio qualsiasi variazione dei parametri di controllo di un algoritmo è lecita. L'impiego per scopi musicali, tuttavia, impone alcune limitazioni alla libertà di scegliere i parametri di controllo. I parametri di controllo a loro volta possono variare nel tempo, divenendo così a loro volta

dei segnali (di controllo). La variazione dei segnali di controllo acquista un diverso significato secondo la scala dei tempi su cui si attua. Se il controllo si attua sulla scala di tempo della (frazione di) nota, parliamo di controllo della dinamica spettrale. Esso infatti viene spesso interpretato in relazione alla variazione a tempo breve dello spettro. Se il controllo si attua nella scala di tempo dell'organizzazione delle note in frasi o entità superiori, parliamo di controllo espressivo. Ad esempio la variazione del pitch delle note rappresenta il controllo espressivo fondamentale nella musica occidentale.

La sintesi elettronica dei suoni sembra offrire una grande libertà nella costruzione dei suoni, sia nella imitazione di quelli naturali, che nella produzione di sonorità originali. Vi sono ormai varie tecniche per riprodurre i suoni desiderati con la fedeltà voluta. Tuttavia scopo della sintesi del suono non è tanto la riproduzione di un segnale voluto, quanto la realizzazione di un generatore suonabile, caratterizzato cioè da una articolazione timbrica paragonabile a quella degli strumenti classici. Il problema si sposta quindi alle possibilità di controllo dell'algoritmo e dell'articolazione timbrica offerte dagli strumenti sintetici.

5.1.1 Obiettivi della sintesi del suono

Tradizionalmente, nella musica occidentale, il suono è caratterizzato da altezza, intensità, durata metrica, timbro e localizzazione spaziale. Sono questi i parametri che il musicista gestisce. La presenza del pitch presuppone un modello di segnale (quasi) periodico. Il pitch è legato alla frequenza del suono e induce nello spettro del segnale una struttura a righe, dove cioè l'energia è concentrata in bande ristrette (righe) a intervalli regolari sullo spettro. Non tutti i suoni hanno altezza definita; in questi casi si parla di spettri continui, caratterizzati da assenza di regolarità nello spettro. Il loudness è legato all'energia del segnale, la durata metrica è alla base della percezione ritmica. Lo spazio e soprattutto il timbro sono i parametri che offrono maggiori possibilità di articolazione nei suoni sintetici o trasformati elettronicamente.

L'esigenza di manipolare questi parametri rimane un aspetto centrale anche nella musica elettronica. Lo scopo della sintesi del suono quindi dovrebbe tendere a realizzare strumenti suonabili più che generatori di segnale, in modo da preservare il rapporto di causa ed effetto che sussiste tra l'azione sul controllo ed il risultato sul suono. Si dovrebbe cioè offrire al musicista uno *strumento* a tutti gli effetti, inteso come entità caratterizzata da certi requisiti di coerenza interna, che si concretizzano in suonabilità, qualità sonora, utilizzabilità all'interno di una partitura.

Lo strumento musicale è importante anche perché, oltre a rappresentare il processo di generazione, può essere visto come astrazione di una classe di suoni caratterizzati da un timbro, un comportamento dinamico, e da certe possibilità espressive. Questo fatto può applicarsi oltre che agli strumenti tradizionali, anche agli strumenti sintetici. Ne risulta che si possono definire classi astratte di suoni sintetici in base al tipo di modello (e algoritmo) usato per la sintesi e per il tipo di controllo offerto al musicista. Una volta, la scelta dell'algoritmo di sintesi avveniva in base alla efficienza computazionale, anche a spese della sua controllabilità. Oggi, con lo sviluppo della tecnologia, questo problema è sempre meno importante.

Acquistano quindi sempre più importanza altri criteri di scelta, tra cui "migliore" metafora per il musicista e "migliore" risultato acustico. Al primo criterio corrisponde il grado di suggestione che l'algoritmo opera sul musicista-compositore; ad esempio la sintesi additiva suggerisce una visione armonica. Al secondo criterio corrisponde l'esigenza di un risultato acustico ben preciso, o di una particolare interfaccia verso l'esecutore; ad esempio la sintesi per modulazione di frequenza può riprodurre facilmente suoni percussivi inarmonici (campane).

Gli strumenti sintetici, al pari degli strumenti classici, sono caratterizzati dal problema dell'apprendimento della tecnica di esecuzione. Si deve infatti imparare con l'esperienza le relazioni tra i

parametri di controllo e il risultato acustico. Queste relazioni spesso non sono intuitive nel controllo a basso livello degli algoritmi e quindi limitano di fatto la versatilità dello strumento. Si può notare d'altra parte che la tendenza attuale è quella di incorporare l'esecutore nello strumento; si cerca cioè di realizzare uno strumento senza problemi di manualità e controllabile con informazioni di alto livello, eventualmente per mezzo di esecutori automatici (sequencer). Nell'ottica di questo approccio devono quindi essere sviluppati sofisticati modelli del controllo timbrico che, a partire da poche e sintetiche informazioni, siano in grado di produrre un ventaglio espressivo paragonabile a quello di un esecutore umano.

Di seguito sono presentati i principali algoritmi di sintesi con riferimento ai criteri di scelta sopra esposti. È tuttavia possibile procedere ad una classificazione degli algoritmi di sintesi basata sull'analisi della loro struttura. Si può infatti notare che la complessità della struttura ha forti riflessi sulla controllabilità sia timbrica che espressiva di un algoritmo. Gli algoritmi definiti da una struttura semplice necessitano di un flusso di segnali di controllo molto articolato, in quanto caratterizzazione ed espressività timbrica devono essere garantiti proprio dai segnali di controllo. Invece gli algoritmi con complessità strutturale elevata garantiscono di base una buona caratterizzazione timbrica e una buona coerenza interna, consentendo quindi un controllo molto più semplificato. Si possono quindi individuare le seguenti classi di algoritmi:

- generazione diretta: di questa classe fanno parte campionamento, sintesi additiva, granulare;
- feed-forward: sottrattiva, modulazioni, distorsione non lineare;
- feed-back: sintesi per modelli fisici

Ad esempio se prendiamo in considerazione uno strumento caratterizzato da un controllo gestuale assai semplice come il pianoforte, si identificano i seguenti requisiti per i segnali di controllo:

- sintesi additiva: supponendo di lavorare con 100 parziali la pressione del tasto attiva 100 involuppi temporali e altrettanti involuppi frequenziali con andamento in funzione della velocità della pressione del tasto.

- sintesi FM: supponendo di lavorare con un algoritmo a 6 operatori la pressione del tasto attiva 6 involuppi temporali e altrettanti involuppi degli indici di modulazione con andamento funzione della velocità della pressione del tasto.

- sintesi per modelli fisici: supponendo di lavorare con un algoritmo martelletto corda, la pressione del tasto fornisce l'unico parametro (la velocità d'impatto del martelletto) all'algoritmo, che provvede autonomamente a generare la sonorità attesa.

È possibile anche un'altra classificazione degli algoritmi di sintesi in base al tipo di modello con cui viene rappresentato il suono. In questo caso si possono distinguere

- *modelli del segnale* che rappresentano il suono che ci arriva all'orecchio, senza far riferimento al meccanismo fisico che sottosta alla produzione del suono. La percezione del suono è un fenomeno complesso, che analizza il segnale sia nel tempo che nella frequenza. Anche i modelli del segnale possono essere divisi in due classi, secondo se possono essere interpretati dall'utente in termini di caratteristiche temporali o spettrali. Possiamo quindi includere nella prima classe il campionamento e la sintesi granulare, mentre la sintesi additiva e sottrattiva, le modulazioni e la distorsione non lineare sono della seconda classe (meglio interpretabili nel dominio della frequenza).
- *modelli della sorgente* che ottengono il segnale acustico come sottoprodotto di un modello di simulazione del meccanismo fisico di produzione del suono. Appartiene a questa categoria la sintesi per modelli fisici.

Va infine ricordato che quando si parla di segnali musicali generalmente si intendono i segnali sonori. Come detto però il risultato acustico che si ottiene da un modello dipende dal controllo che si effettua sui parametri del modello stesso. In molti casi questi parametri sono tempo varianti e si evolvono durante lo sviluppo del singolo suono. Sono cioè essi stessi dei segnali, chiamati appunto di controllo, che però si differenziano dai segnali audio perché si evolvono più lentamente. Inoltre essi vengono percepiti seguendo la loro evoluzione temporale e non analizzandoli in frequenza, come accade per i segnali audio. Nel seguito verranno esposti i principali algoritmi di sintesi dei segnali audio. Talvolta essi sono utili anche per i segnali di controllo. Alla fine verrà discussa la problematica della generazione dei segnali di controllo.

5.2 Metodi di generazione diretta

In questa categoria troviamo i metodi che generano direttamente il segnale attraverso un'unico modello o più modelli che però non si influenzano reciprocamente, nel senso che al più si somma alla fine il loro risultato.

5.2.1 Generatori di forme d'onda

Oscillatori numerici

La caratteristica di molti suoni musicali è di essere quasi periodici o armonici. È questa proprietà che determina la sensazione di altezza di un suono. Il più semplice metodo di sintesi consiste nel produrre un segnale periodico mediante la continua ripetizione di una certa forma d'onda. Un algoritmo che realizza questo metodo si chiama oscillatore. L'oscillatore più diffuso è quello a forma d'onda tabulata (*table look-up oscillator*). In questo caso la forma d'onda è memorizzata in una tabella in punti equispaziati. Per generare una forma d'onda periodica, basta leggere ripetutamente la tabella mandando i suoi campioni uno dopo l'altro in uscita. Se F_s è la frequenza di campionamento e L è la lunghezza della tabella, la frequenza f del suono periodico risulta $f = F_s/L$. Se si volesse un suono con la stessa forma d'onda ma di frequenza diversa, occorrerebbe una tabella contenente la stessa forma d'onda ma rappresentata con un numero diverso di valori. Si vorrebbe quindi una forma d'onda continua da cui prelevare di volta in volta il valore all'ascissa desiderata. A questo scopo si ricorre a tabelle contenenti la forma d'onda in (molti) punti equispaziati e poi prelevando di volta in volta il valore più opportuno o mediante interpolazione tra i due punti adiacenti o usando il valore di ascissa più prossima a quella desiderata (interpolazione di ordine zero). Naturalmente più fitti sono i punti, migliore è l'approssimazione. Si usano tipicamente tabelle da 256 a 4096 punti. In questo modo l'oscillatore ricampiona la tabella per generare un suono di differente frequenza.

La distanza (in numero di campioni della tabella) fra due campioni della tabella prelevati in istanti successivi si chiama *SI* (*sampling increment*) ed è proporzionale alla frequenza f del suono prodotto. Risulta infatti

$$f = \frac{SI \cdot F_s}{L}$$

L'algoritmo che implementa l'oscillatore digitale è

```
phi = mod(phi + SI, L)
s = A * tab[phi]
```

dove ϕ rappresenta la fase istantanea (ascissa) di lettura nella tabella, A è un parametro indicante l'ampiezza del segnale, s è il segnale di uscita. La funzione $\text{mod}(x, L)$ calcola il resto della

divisione del primo operando (\times) per il secondo (L) e serve per ricominciare la lettura della tabella al punto opportuno, dopo aver scorso tutta la forma d'onda.

Se il passo di lettura SI e' maggiore di uno, puo' succedere che le frequenze delle componenti piu' alte siano maggiori della frequenza di Nyquist, dando luogo a foldover. Per evitare questo fenomeno, bisogna limitare la banda del segnale memorizzato. Se invece il passo e' minore di uno, come avviene spesso per i segnali di controllo, involuppi di ampiezza etc., allora il problema non si pone in quanto la banda e' gia' sufficientemente limitata.

Normalmente si richiede che l'ampiezza del suono vari in modo continuo seguendo un involuppo d'ampiezza. Questo si puo' ottenere variando istante per istante il parametro A visto precedentemente. Similmente si puo' variare istantaneamente la frequenza del suono variando il parametro SI . Si ottiene cosi' l'oscillatore ad ampiezza e frequenza variabile descritto dalle seguenti relazioni:

$$\begin{aligned}\phi(n) &= (\phi(n-1) + SI(n))_{mod L} \\ s(n) &= A(n) \cdot tab[\phi(n)]\end{aligned}$$

La prima formula effettua l'integrazione discreta della frequenza, per trovare la fase istantanea.

Il controllo dell'ampiezza serve per realizzare un appropriato involuppo del suono. Inoltre si puo' ad esempio realizzare l'effetto del tremolo, che consiste in una variazione periodica o quasi periodica dell'ampiezza attorno al valore medio. Il controllo della frequenza dell'oscillatore e' utile nella produzione di variazioni di pitch come nel portamento, inflessione della frequenza del suono come puo' avvenire all'inizio di una nota che parte con una frequenza leggermente inferiore raggiungendo poi ed eventualmente sorpassando la frequenza desiderata oppure la diminuzione di frequenza che talvolta si verifica alla fine di una nota. Spesso occorrono opportune traiettorie della frequenza per collegare in modo non brusco note di una stessa frase melodica o per produrre ornamenti come trilli, mordenti e acciaccature. Infine il controllo della frequenza e' utile per produrre il vibrato ed altre modulazioni di frequenza. In figura 5.1 e' riportato il simbolo dell'oscillatore a forma d'onda fissa con ampiezza e frequenza variabili.

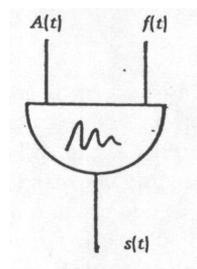


Figura 5.1: Simbolo dell'oscillatore a forma d'onda fissa con ampiezza e frequenza variabili

I suoni ottenibili dall'oscillatore numerico sono piuttosto meccanici e ripetitivi. Esso pertanto viene usato come blocco base per costruire algoritmi piu' complessi o per produrre segnali di controllo ad andamento voluto.

Lo stesso procedimento puo' essere usato per generare forme d'onda non ripetitive. In questo caso la forma d'onda viene sempre memorizzata in una tabella. Per generare una forma d'onda di durata d si puo' utilizzare poi l'algoritmo precedente con passo di campionamento

$$SI = \frac{L}{d \cdot F_s}$$

Si puo' generalizzare il procedimento generando forme d'onda date per punti (coppie di ascisse e ordinate) e interpolando via via fra i punti adiacenti. Ad esempio un involuppo d'ampiezza puo' essere descritto da una spezzata composta da vari punti connessi da linee rette. Se si vuole cambiare la durata dell'involuppo, e' bene modificare poco le durate dei tratti corrispondenti all'attacco e decadimento del suono, mentre si puo' variare di piu' il tratto di regime. In questo modo si avranno differenti passi di lettura della tabella o distanza tra le ascisse dei punti generati a seconda di quale porzione dell'involuppo si sta generando.

Esempio di sintesi in Matlab

Molti linguaggi orientati alla sintesi del suono (come il noto *Csound*), adottano degli accorgimenti per ottimizzare il carico computazionale al fine di ridurre i tempi di attesa dell'elaborazione. Attualmente questi tipi di linguaggi sono in grado di rispondere in tempo reale agli ingressi di controllo, attraverso una gestione dei segnali di controllo al *frame rate*: un frame e' una finestra temporale di durata prestabilita (valori tipici della durata di un frame vanno dai 5 ai 50 msec). L'ipotesi che generalmente si fa sui segnali di controllo e' che essi cambino ad una frequenza molto inferiore della frequenza di campionamento audio e possano essere considerati costanti all'interno di un frame.

Il linguaggio che si e' scelto per illustrare i principali algoritmi di sintesi (MATLAB) non e' un linguaggio orientato alla sintesi audio e non consente il controllo degli algoritmi in tempo reale. Si manterra' tuttavia la divisione di due assi temporali (*frame rate* / *audio rate*), in quanto elemento comune a tutti i linguaggi piu' diffusi.

Ogni sessione di simulazione necessita della definizione di alcuni parametri globali, come le frequenze di campionamento e di controllo, che saranno visti dalle funzioni che realizzano gli algoritmi. Un esempio di tale sezione di definizione e' il seguente

```
global Fs; %sample rate
Fs=22050
ControlW=0.01 % control window (in sec): 10 ms
global SpF; %samples per Frame
SpF=round(Fs*ControlW)
Fc=Fs/SpF %control rate
```

Un tipico script MATLAB per l'esecuzione di un algoritmo di sintesi prevede tre fasi: un'instatazione con le definizioni, una fase di generazione dei segnali di controllo e, infine, una fase di generazione del segnale audio. Nello scrivere una funzione di sintesi si segue la convenzione per cui la durata (in secondi) dei segnali di controllo determina la durata del suono generato dall'oscillatore. Supponendo che la funzione `sinosc(t0, a, f, ph0)` realizzi un oscillatore sinusoidale (con t_0 istante di inizio e a , f , e $ph0$ rispettivamente ampiezza, frequenza e fase iniziale della sinusoide), le istruzioni seguenti generano un sinusoide di durata 2 secondi, ampiezza unitaria e frequenza 50 Hz (con riferimento alle definizioni globali esemplificate).

```
a=ones(1,200); %1secondo, 100 frames
f=50*ones(1,200);
s=sinosc(0,a,f,0);
```

Si vuole realizzare ora l'oscillatore sinusoidale controllato in ampiezza e a frequenza imposta al *frame rate*. Si noti come e' stata curata la continuita' della fase per evitare distorsioni della forma d'onda a fronte di variazioni istantanee della frequenza. Si noti inoltre che per variazioni a scalino della frequenza da $f(i)$ a $f(i+1)$, nel frame di transizione la frequenza e' interpolata linearmente (con derivata costante pari a $\frac{f(i+1)-f(i)}{Spf}$).

```

function s = sinosc(t0,a,f,phi0);
% oscillatore sinusoidale
% a e f sono al control rate
% t0 istante iniziale in sec;

global SpF; %samples per frame
global Fs; %sampling rate
Fc=Fs/SpF; %control rate

nframes=length(a); %numero di frames
if (length(f)==1) f=f*ones(1,nframes); end
if (length(f)~=nframes)
    error('f e a devono avere lunghezza uguale');
end

s=zeros(1,nframes*SpF);
t=0:(1/Fs):((nframes*SpF)/Fs);

lastfreq=f(1);
lastphase=0;
for (i=1:nframes)
    taux=0:(1/Fs):((SpF-1)/Fs);
    phase=lastphase+lastfreq*2*pi.*taux+(1/SpF)*pi*(f(i)-lastfreq).*(taux.^2);
    s(((i-1)*SpF+1):i*SpF)=a(i).*cos(phase);
    lastfreq=f(i);
    lastphase=phase(SpF);
end

s=[zeros(1,(t0*Fs-1)) s];

```

In ogni linguaggio di sintesi subito dopo l'oscillatore sinusoidale si incontra, per importanza, la famiglia dei generatori di funzioni di controllo. Nel seguito e' riportata la realizzazione di un semplice generatore di involuppo a segmenti di retta (si noti, tuttavia, che la funzione di interpolazione del MATLAB permette di avere facilmente interpolazioni cubiche o *spline*). A fronte della descrizione della forma di involuppo con istanti temporali (in secondi) e la relativa ampiezza, esso genera la funzione al frame rate.

```

function env = envgen(t,a,method);

global SpF; %samples per frame
global Fs; %sampling rate

if (nargin<3)
    method='linear';
end

frt=floor(t*Fs/SpF+1); %tempi espressi in numero del frame
nframes=frt(length(frt)); %numero totale frames
env=interp1(frt,a,[1:nframes],method);

```

L'uso combinato dei due generatori appena visti permette di creare agevolmente suoni con involuppo di ampiezza e curve di altezza molto articolati. Nell'esempio che segue, l'involuppo e' un

esempio di inviluppo *ADSR*: l'andamento temporale dell' ampiezza di un suono e' suddiviso nelle quattro fasi di *Attack, Decay, Sustain e Release* .

```
f=envgen([0,.2,1,2],[200,250,250,200],'cubic'); % curva di pitch
% curva di pitch con vibrato:
f=f+max(f)*0.05*sin(2*pi*5*(SpF/Fs)*[0:length(f)-1]).*hanning(length(f))';
a=envgen([0,.2,1,1.5,2],[0,1,.8,.5,0],'linear'); %ADSR

% genera segnali audio
s=sinosc(0,a,f,0);
```

In fig. 5.2 sono illustrati i segnali di controllo e il segnale generati.

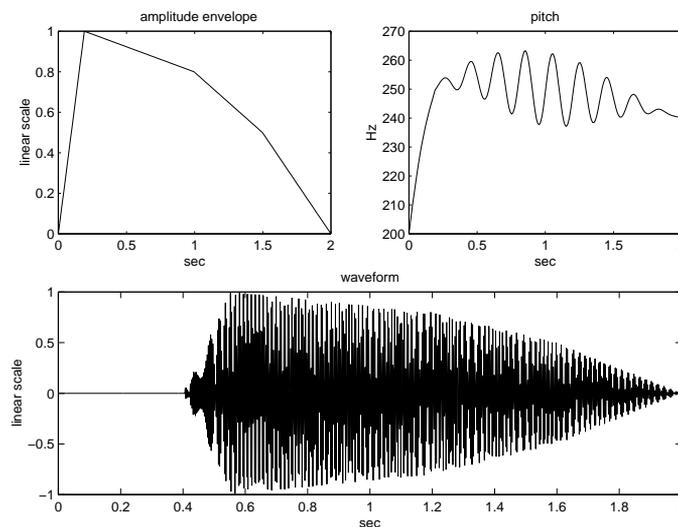


Figura 5.2: Segnali di controllo e segnale ad audio rate

Generatori ricorsivi di segnali sinusoidali

I segnali sinusoidali possono essere generati, oltre che con il metodo dell'oscillatore a tabella, anche con metodi ricorsivi. Un primo metodo si basa sul risonatore numerico, costituito da un filtro del secondo ordine con i poli (complessi coniugati) sul cerchio di raggio unitario. Esso e' dato dall'equazione ricorrente

$$y(n+1) = 2\cos(\omega)y(n) - y^2(n-1) \quad (5.1)$$

dove $\omega = 2\pi f/F_s$. Con condizioni iniziali $y(0) = 1$ e $y(-1) = \cos \omega$ il generatore produce $y(n) = \cos n\omega$; con $y(0) = 0$ e $y(-1) = -\sin \omega$ il generatore produce $y(n) = \sin n\omega$. In generale se $y(0) = \cos \phi$ e $y(-1) = \cos(\phi - \omega)$ il generatore produce $y(n) = \cos(n\omega + \phi)$. Questa proprieta' puo' anche essere verificata ricordando la relazione trigonometrica $\cos\omega \cdot \cos\phi = 0.5[\cos(\phi + \omega) + \cos(\phi - \omega)]$.

Un'altro metodo si basa sulla forma accoppiata decritta dalle equazioni

$$\begin{aligned} x(n+1) &= \cos\omega \cdot x(n) - \sin\omega \cdot y(n) \\ y(n+1) &= \sin\omega \cdot x(n) + \cos\omega \cdot y(n) \end{aligned}$$

Con $x(0) = 1$ e $y(0) = 0$ si ha $x(n) = \cos(n\omega)$ e $y(n) = \sin(n\omega)$; vengono generati contemporaneamente un seno e un coseno. Questa proprieta' puo' essere verificata considerando che se si definisce una

variabile complessa $w(n) = x(n) + jy(n) = \exp(jn\omega)$, risulta $w(n+1) = \exp(j\omega) \cdot w(n)$. Prendendo la parte reale e immaginaria di questa relazione risulta la forma accoppiata.

In generale entrambi metodi hanno il problema che la quantizzazione dei coefficienti puo' causare instabilita' numerica e cioe' i poli non sono esattamente sul cerchio unitario. Le forme d'onda generate allora o tenderanno a smorzarsi o a crescere indefinitamente. A questo scopo e' opportuno periodicamente reinizializzare la ricorsione.

Generatori di rumori

Generazione di numeri aleatori Per generare un rumore si ricorre ai generatori di numeri pseudo-casuali. Ci sono molti metodi e nessuno e' soddisfacente sotto tutti gli aspetti. Il metodo piu' diffuso si chiama congruenziale lineare e puo' generare sequenze piuttosto lunghe di numeri aleatori prima di ripetersi periodicamente. Dato un valore iniziale (seme) $I(0)$ nell'intervallo $0 \leq I(0) < m$, l'algoritmo di generazione si basa sulla ricorrenza

$$\begin{aligned} I(n) &= [aI(n-1) + c] \bmod(m) \\ s(n) &= I(n)/m \end{aligned}$$

dove a e c sono due costanti che devono essere scelte accuratamente in relazione al valore di m , per riuscire ad avere la sequenza di lunghezza massima. I numeri generati $s(n)$ sono uniformemente distribuiti nell'intervallo $0 \leq s(n) < 1$. Su questo intervallo la densita' di probabilita' e' piatta. Pertanto la media vale $E[u] = 1/2$ e la varianza $\sigma_u^2 = 1/12$. Per avere una sequenza a media nulla si fa $u(n) = s(n) - 0.5$. Questa sequenza corrisponde ad un rumore bianco in quanto i numeri generati possono essere considerati mutualmente indipendenti e la densita' spettrale di potenza e' data da $S(f) = \sigma_u^2$. Pertanto, essendo $S(f)$ piatto, la sequenza contiene tutte le frequenze in ugual misura e presenta parimenti variazioni lente e veloci.

Generazione di rumori a bassa frequenza Se si desidera che la sequenza vari piu' lentamente, si puo' generare un nuovo numero aleatorio ogni d campioni e mantenendo il precedente nell'intervallo (*holder*) o facendo interpolazione lineare tra i due valori generati. In questo caso lo spettro di potenza e' dato da:

$$S(f) = |H(f)|^2 \frac{\sigma_u^2}{d}$$

con

$$|H(f)| = \left| \frac{\sin(\pi f d / F_s)}{\sin(\pi f / F_s)} \right|$$

nel caso dell'holder e

$$|H(f)| = \frac{1}{d} \left[\frac{\sin(\pi f d / F_s)}{\sin(\pi f / F_s)} \right]^2$$

nel caso di interpolazione lineare.

Generazione di rumori $1/f$ Un rumore $1/f$, chiamato anche rumore rosa, e' caratterizzato da uno spettro di potenza $S(f)$ che decresce in frequenza secondo un andamento proporzionale a $1/f$

$$S(f) = \frac{A}{f} \quad (5.2)$$

In genere, per evitare un valore infinito a $f = 0$, si considera questa espressione valida per $f \geq f_{min}$, dove f_{min} e' la minima frequenza desiderata. Lo spettro (5.2) e' caratterizzato da un decadimento di 3 dB per ottava, cioe' quando la frequenza raddoppia, lo spettro di potenza si dimezza. L'ammontare di potenza contenuta in un intervallo di frequenza $[f_1, f_2]$ e'

$$\int_{f_1}^{f_2} S(f)df = A \ln\left(\frac{f_1}{f_2}\right)$$

Questo implica che l'ammontare di potenza contenuta in ogni intervallo di ottava e' sempre la stessa. Il rumore $1/f$ e' presente in molti fenomeni naturali ed e' legato ai fenomeni frattali. Nell'audio e' conosciuto come rumore rosa, per differenziarlo dal rumore bianco. Esso rappresenta l'equivalente psicoacustico del rumore bianco, in quanto contiene all'incirca la stessa potenza per ogni banda critica. In senso fisico esso dipende da processi che si evolvono su differenti scale temporali. Un modello per generare rumore $1/f$ consiste nella somma di vari rumori bianchi, ciascuno filtrato attraverso un filtro passa-basso del primo ordine e con costante di tempo via via crescente in progressione geometrica. Una variante proposta da Voss consiste nel prendere la media di vari generatori y_i a tenuta di numeri aleatori con periodo di rinnovamento $d_i = 2^i$; cioe'

$$y(n) = \frac{1}{M} \sum_{i=1}^M y_i(n) \quad (5.3)$$

Lo spettro di potenza di (5.3) non ha esattamente un andamento del tipo $1/f$, ma lo approssima per frequenze $f \geq F_s/2^M$.

5.2.2 Campionamento

Trovare un modello matematico che imiti fedelmente un suono reale e' un compito estremamente difficile. Se pero' esiste un suono di riferimento, allora e' sempre possibile riprodurlo dopo averlo registrato digitalmente mediante campionamento (*sampling*). Tale metodo, anche se semplice nei suoi principi, e' molto usato negli strumenti musicali digitali e, appunto, nei campionatori. I campionatori infatti memorizzano una grande quantita' di esempi di suoni completi, usualmente prodotti da strumenti musicali reali. Quando si vuole sintetizzare un suono, basta scegliere uno dei suoni del repertorio memorizzati e riprodurlo direttamente. Ne risulta quindi una alta efficienza computazionale e una grande fedelta' al suono originale.

Le possibilita' di modificazione sono piuttosto ridotte e sono spesso legate alla metafora del registratore a nastro o moviola. La modificazione piu' frequente consiste nel cambiare la frequenza del suono, variando la frequenza di lettura dei campioni. Non sono consigliabili grandi variazioni di frequenza, in quanto la compressione o espansione temporale di una forma d'onda produce un cambiamento inverso della scale delle frequenze e quindi un'espansione o compressione dello spettro. Tale fatto tende a produrre un risultato innaturale dal punto di vista timbrico, esattamente come accade se viene variata la velocita' di lettura di un nastro magnetico. E' pertanto necessario limitare le variazioni a pochi semitoni ed avere quindi molti suoni campionati distribuiti lungo la scala musicale. Speciale cura va posta in questo caso per non avere suoni adiacenti troppo diversi. Con un insieme di suoni (ad esempio tre per ottava) e con la variazione di lettura dei campioni e' quindi possibile riprodurre tutta la gamma di altezze desiderate.

Spesso si vuole inoltre variare il suono anche in funzione di altri parametri, ad esempio l'intensita'. Per ottenere una variazione di intensita' non basta infatti cambiare l'ampiezza del suono, ma bisogna anche modificare timbricamente il suono. Tipicamente i suoni piu' intensi sono caratterizzati

da un attacco piu' rapido e da una maggiore estensione dello spettro. In tal caso o si utilizza un unico prototipo (ad esempio registrato fortissimo) e poi lo si trasforma (ad esempio mediante filtraggio) per ottenere le altre intensita', o si ricorre ad insiemi diversi di note registrate per differenti valori del parametro (ad esempio con dinamica fortissimo, mezzo forte, pianissimo) e poi si procede a creare le varie sfumature con interpolazioni e/o ulteriori trasformazioni. In definitiva questa tecnica e' caratterizzata da alta efficienza computazionale e alta qualita' imitativa, ma bassa flessibilita' per i suoni non inizialmente previsti nel repertorio o non facilmente riconducibili a esso con semplici trasformazioni.

Per maggior efficienza nell'uso della memoria, spesso si ricorre a memorizzare solo parte del regime stazionario del suono e a ripeterlo (*looping*) nella sintesi. Naturalmente la ripetizione non deve essere di un segmento troppo breve per evitare un carattere troppo statico del suono. Ad esempio per allungare la durata di un suono, dopo che e' passato l'attacco si puo' ripetere ciclicamente la parte individuata finche' non si vuole terminare il suono. A quel punto si emette la parte finale del suono memorizzato. Per creare un ciclo senza artefatti, bisogna porre molta cura nello scegliere i punti di inizio e fine del ciclo. In genere si sceglie un numero intero di periodi inizianti con valore nullo in modo da non avere discontinuita' ne' di ampiezza ne' di fase. Queste discontinuita' infatti sono fastidiose all'ascolto.

Spesso si individuano nel regime alcuni brevi tratti significativi e nella sintesi si procede ad una interpolazione (*cross-fade*) tra i successivi tratti. In questo modo l'evoluzione temporale lungo la durata del suono puo' essere meglio controllata.

In molti casi tale tecnica viene presentata come un mezzo per riprodurre suoni naturali ed e' valutata facendo riferimento agli strumenti originali. Per questo essa e' molto usata nelle tastiere commerciali per produrre suoni imitativi degli strumenti meccanici, come ad esempio organo o piano elettronici. Naturalmente il metodo di campionamento non puo' realizzare tutte le possibilita' espressive degli strumenti originali. D'altra parte si puo' notare che i suoni memorizzati possono essere sintetici o derivare da modificazioni di altri suoni. Questo amplia le possibili applicazioni del metodo. Dal punto di vista della storia della musica, questo metodo rappresenta una versione attualizzata della Musica Concreta. Questo tipo di musica, nata a Parigi nel 1950 per opera soprattutto di Pierre Schaefer, inizio' ad usare come materiale sonoro delle composizioni musicali suoni di qualsiasi tipo registrati da microfono e poi eventualmente manipolati.

5.2.3 Sintesi additiva

Per avere un suono con un andamento complesso e variabile si possono usare contemporaneamente piu' elementi semplici. Per esempio un'orchestra produce una sonorita' piena e variabile mediante una sovrapposizione di molti strumenti.

Nella sintesi additiva, suoni complessi sono prodotti mediante la sovrapposizione di suoni elementari, spesso sinusoidali. In certe condizioni (ad esempio se le frequenze sono multiple tra loro) i suoni costituenti si fondono insieme e il risultato e' percepito come un unico suono. Questo procedimento e' usato anche in alcuni strumenti musicali tradizionali. Ad esempio nell'organo le canne producono suoni relativamente semplici. Per ottenere suoni piu' ricchi, sono usate piu' canne che suonano contemporaneamente a differenti altezze, tipicamente multiple della fondamentale.

Quando si analizza un suono reale quasi periodico prodotto da uno strumento musicale, si osserva che l'energia spettrale si concentra attorno a poche righe strette. Queste righe corrispondono a sinusoidi, chiamate parziali, di frequenza all'incirca multipla della fondamentale. L'ampiezza di ciascuna parziale non e' proporzionalmente costante nel tempo rispetto alla fondamentale, ma varia secondo un suo andamento. Ad esempio nell'attacco, sono significative alcune parziali, che poi diventano trascurabili nel regime. In generale le parziali sono componenti di uno spettro a frequenza arbitra-

ria. Le parziali cioè possono essere o non essere armoniche (multipli interi) di una fondamentale. In un suono inarmonico (come quello di una campana) le parziali non sono in rapporti interi con la fondamentale.

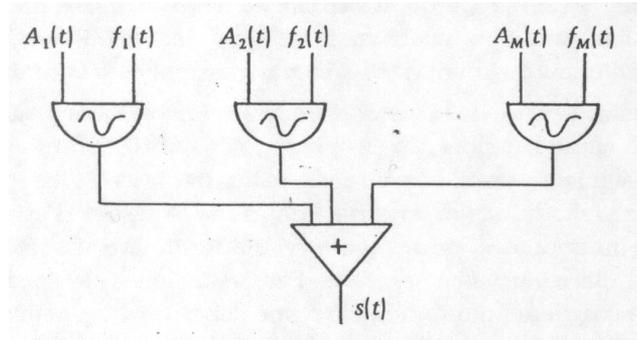


Figura 5.3: Somma di oscillatori sinusoidali con ampiezza e frequenza tempo varianti

La formula della sintesi additiva (sinusoidale) e' pertanto:

$$s(n) = \sum_k A_k(n) \cdot \sin\left(2\pi \frac{f_k(n)}{F_s} n + \phi_k\right)$$

Essa si puo' realizzare mediante la somma di oscillatori sinusoidali, visti precedentemente, con ampiezza $A_k(n)$ e frequenza $f_k(n)$ tempo varianti (fig. 5.3). Nei i suoni armonici le frequenze delle parziali sono caratterizzate da $f_k \simeq k f_1$. Come esempio di suono inarmonico, si puo' ottenere una imitazione dei suoni di campana, mediante la somma di componenti sinusoidali a frequenze non legate armonicamente corrispondenti ai modi della campana, con un inviluppo d'ampiezza esponenziale decrescente, piu' lungo per le componenti piu' gravi che si smorzano piu' lentamente. Per creare i battimenti si possono controllare ulteriormente l'ampiezza o aggiungere delle componenti di frequenza prossima ad alcune di quelle esistenti.

Se scegliamo i parametri di controllo desunti dall'analisi spettrale di un suono naturale, questa tecnica di sintesi consente un buon grado di riproduzione. La sintesi additiva fornisce anche una buona riproduzione dei suoni non periodici o inarmonici in cui l'energia sia concentrata in poche righe spettrali. Molto meno adatta e' invece per la parte di suono che e' distribuita nello spettro, come le componenti rumorose.

La sintesi additiva e' un metodo di grande generalita' ma ha il problema che ha bisogno di un gran numero di parametri di controllo per ciascuna nota. Devono essere specificate due funzioni di controllo per ciascuna parziale (ampiezza e frequenza). Normalmente le funzioni sono diverse per ogni nota e dipendono dalla durata e intensita' della nota. Essa pertanto e' spesso usata per la sintesi basata sull'analisi. Infatti, come si vedra', le trasformazioni fatte sui parametri della sintesi additiva (come cambiamento della scala dei tempi) sono particolarmente robuste percettivamente. Inoltre e' utile per familiarizzare i musicisti con le caratteristiche del suono e la rappresentazione in frequenza.

In conclusione si puo' anche osservare che questo metodo, nato per simulare suoni reali, e' diventato per alcuni musicisti, il fondamento metaforico per una metodologia compositiva basata sull'espansione della scala temporale e la reinterpretazione dello spettro come struttura armonica.

Esempio di sintesi additiva in Matlab

Il generatore sinusoidale controllato in ampiezza e frequenza costituisce l'unita' fondamentale degli algoritmi di sintesi additiva. Si pu' utilizzare l'oscillatore realizzato nell'esempio precedente per sperimentare il fenomeno dei battimenti. L'esperimento prevede due oscillatori di cui uno a frequenza costante e l'altro a frequenza crescente (o decrescente):

```
% genera segnali di controllo
f1=envgen([0,20],[200,200]); %sinusoide fissa
%sinusoide a frequenza crescente:
f2=envgen([0,1,5,10,15,20],[200,200,205,220,270,300]);
a=envgen([0,5,10,15,20],[1,1,1,1,1]);

% genera segnali audio
s=sinosc(0,a,f1,0)+sinosc(0,a,f2,0);
```

La fig. 5.4 mostra i segnali di controllo della frequenza e l'effetto di involuppo di ampiezza caratteristico dei battimenti del primo ordine.

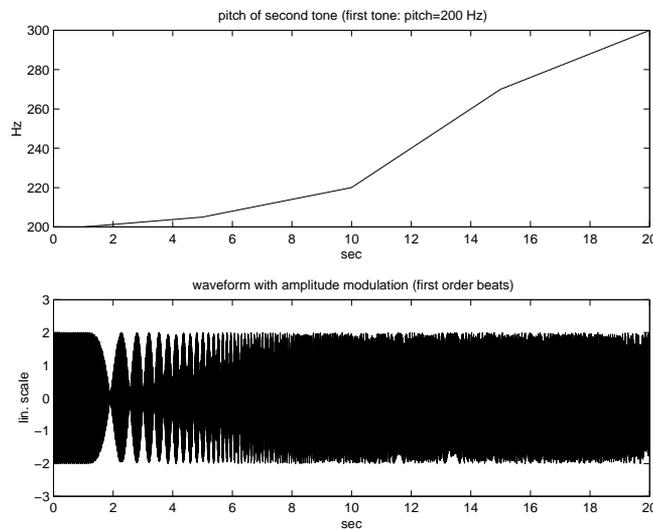


Figura 5.4: Battimenti del primo ordine

La sintesi additiva permette di ottenere suoni di sintesi di alta qualita' se i parametri di frequenza e ampiezza delle parziali vengono estratti dalla versione campionata del suono da riprodurre attraverso una analisi di tipo *Short Time Fourier Transform (STFT)*. Come e' noto, questo tipo di analisi permette di costruire le traiettorie temporali (al frame rate) di frequenza e ampiezza delle singole parziali che costituiscono il suono. In figura 5.5 si puo' osservare il risultato di un analisi di questo tipo per una nota di sassofono.

A partire dalle due matrici contenenti i dati ora visti, e' facile ottenere la risintesi del suono di partenza con sintesi additiva:

```
% genera segnali di controllo
readsan; %legge file di analisi: sinan_freqs e sinan_amps

% genera segnali audio
```

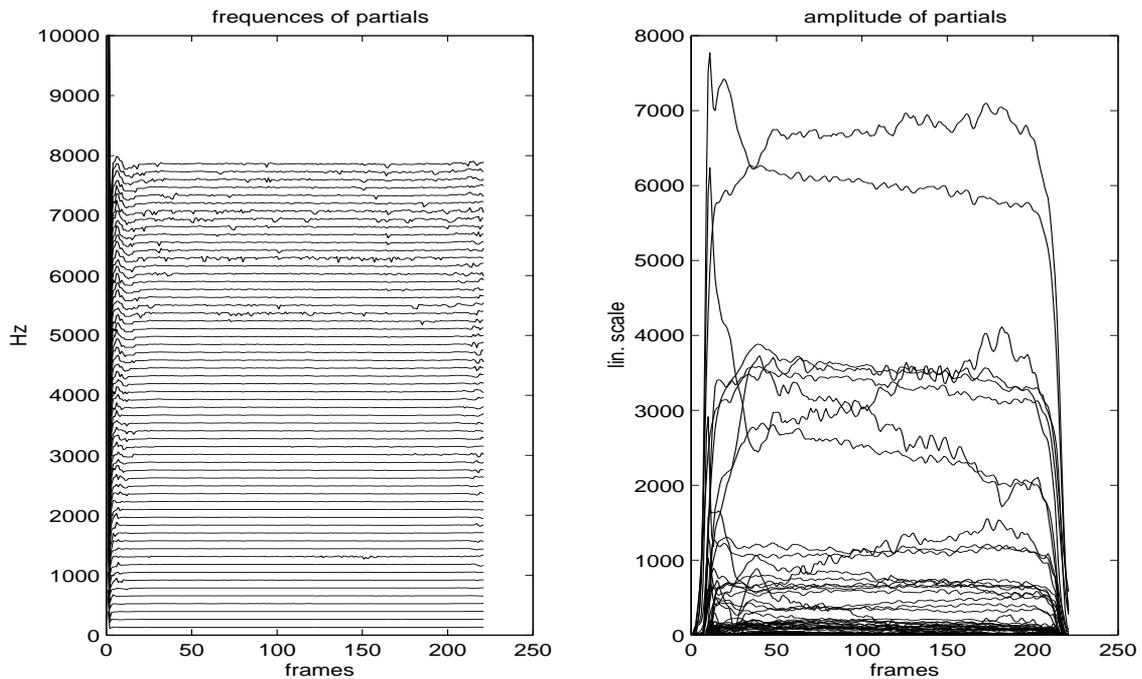


Figura 5.5: Analisi STFT di una nota di sassofono: frequenza (a sinistra) e ampiezza (a destra) delle parziali nel tempo

```

nparz=size(sinan_amps,1);
%genera la fondamentale nel tempo
s=sinosc(0.5,sinan_amps(1,:),sinan_freqs(1,:),0);
for (i=2:nparz)
    %genera le parziali superiori
    s=s+sinosc(0.5,sinan_amps(i,:),sinan_freqs(i,:),0);
end

```

5.2.4 Sintesi granulare

La sintesi granulare condivide con la sintesi additiva l'idea di comporre suoni complessi a partire da suoni più semplici. Mentre la sintesi additiva si basa sulla sovrapposizione temporale di sinusoidi, la sintesi granulare invece si basa sulla successione di forme d'onda di breve durata (tipicamente da 1 a 100 msec) chiamate *grani*. Da questo punto di vista un grano è un breve evento acustico la cui durata è prossima alle soglie di discriminazione della durata, frequenza e intensità nella percezione uditiva. È un po' come nel cinema dove la successione veloce di immagini statiche, produce la sensazione di movimento. Questa idea base si articola poi in due casi principali a seconda della forma d'onda del grano.

Granulazione di suoni

Nel primo, forme d'onda complesse, prese da suoni reali [Truax] o descritte come spettri, si susseguono in parte sovrapponendosi nel metodo chiamato *Overlap and Add (OLA)*. Si possono così riprodurre fedelmente suoni che modificarli nelle caratteristiche dinamiche. Come nella sintesi additiva era importante la coordinazione nella scelta delle frequenze, così qui è importante curare

l'allineamento temporale dei grani, allo scopo di evitare fenomeni di discontinuita' di fase, che producono effetti acustici poco gradevoli. Questo rende spesso il controllo difficile. Un esempio d'uso e' nella sintesi della componente stocastica del segnale, come descritto da X. Serra[1996]; in questo caso viene infatti solo controllata l'evoluzione dell'involuppo spettrale. A questo scopo per ogni frame si ricorre alla trasformata di Fourier inversa, mediante FFT, di uno spettro il cui modulo e' definito dall'involuppo spettrale e la fase viene creata da un generatore di numeri casuali. Ogni frame viene poi moltiplicato per una finestra prima di fare l'OverLap-Add, cioe' la somma dei vari frames con parziale sovrapposizione temporale. Si puo' usare questo approccio anche come metodo di trasformazione di suoni campionati (granulazione di suoni). In questo caso i grani vengono costruiti prelevando piccole parti di un suono, registrato precedentemente o acquisito direttamente da un convertitore D/A, e poi applicando ad ognuna un involuppo di ampiezza. Questi grani poi vengono emessi con ordine scelto dal compositore, ad esempio con velocita' variabile o mescolandoli con ordine casuale. I grani possono anche essere scelti da suoni diversi e poi emessi in modo interlacciato, creando quindi tessiture intermedie (fig. 5.6).

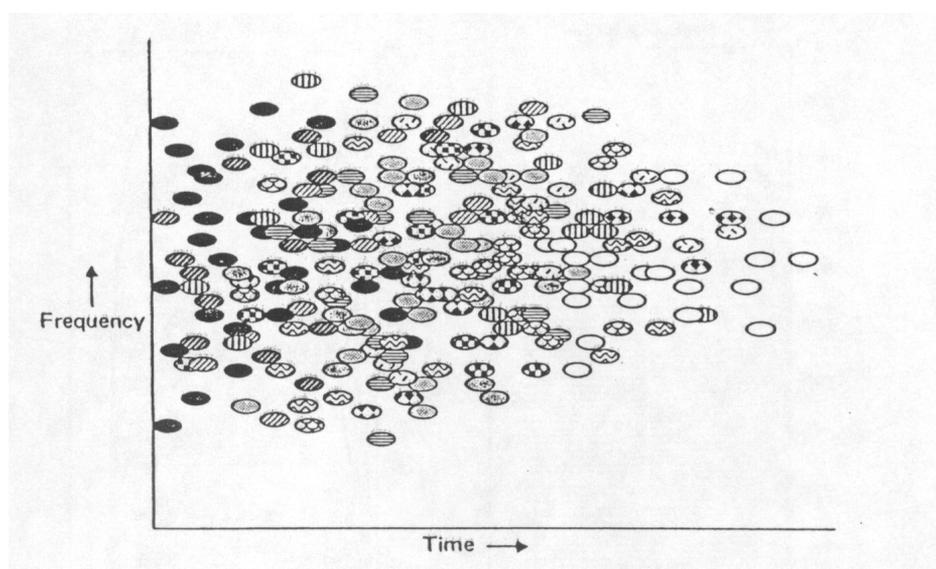


Figura 5.6: rappresentazione grafica di sintesi granulare con grani ricavati da sorgenti diverse e mescolati in modo aleatorio che evolve nel tempo [da Roads 91]

Sintesi

Nel secondo tipo invece si usano come grani funzioni gaussiane (o in generale di tipo passabasso) modulate in frequenza, in modo da localizzare l'energia nel piano tempo-frequenza. Abbiamo qui invece un'analogia con il mosaico, dove l'analogo del grano e' la singola tessera monocromatica e la giustapposizione di tessere di colori diversi fornisce un'immagine complessa. In questo caso la forma d'onda dell' i -esimo grano e' data da

$$g_i(n) = w_i(n) \cdot \cos\left(2\pi \frac{f_i}{F_s} n + \phi_i\right)$$

dove $w_i(n)$ e' una finestra di durata N_i campioni. La formula di sintesi e' data da

$$s(n) = \sum_i a_i \cdot g_i(n - n_i)$$

dove a_i e' l'eventuale coefficiente di ampiezza del grano i -esimo e n_i e' il suo riferimento temporale. Ogni grano da' quindi un contributo di energia concentrato nel piano tempo-frequenza attorno al punto (n_i, f_i) .

Quando i grani sono collocati regolarmente su una griglia nel piano tempo frequenza, essa diventa una realizzazione della sintesi da analisi tempo-frequenza tipo STFT. In questo caso l'analogia e' l'immagine a colori sullo schermo di un computer, composta da una griglia di pixel di tre colori. Quando invece i grani sono sincroni con il periodo del segnale, si ha la cosiddetta sintesi granulare sincrona con il periodo, che fa riferimento alla sintesi sottrattiva come filtraggio di un segnale quasi periodico. Infatti ogni grano puo' essere interpretato come la risposta all'impulso di un filtro FIR e quindi il risultato puo' essere interpretato come un treno periodico di impulsi che eccita un banco di filtri FIR tempo varianti. Questa interpretazione fornisce anche i criteri per la scelta delle forme d'onda dei grani.

Il caso piu' importante e classico di sintesi granulare e' quando invece i grani semplici sono distribuiti in modo irregolare[Roads 1991] (*asynchronous granular synthesis*). Per esempio distribuendo casualmente i grani dentro una maschera che delimita una particolare regione nello spazio tempo-frequenza-ampiezza si ottiene come risultato una nuvola di microsuooni o tessitura (*texture*) musicale che varia nel tempo (fig. 5.7). Si puo' inoltre controllare la densita' dei grani dentro la maschera. Vengono cosi' modellati suoni articolati dove non interessa controllare esattamente la microstruttura. Si evitano cosi' i problemi del controllo dettagliato delle caratteristiche temporali dei grani. La durata dei grani influenza la tessitura sonora: durate brevi danno un carattere scoppiettante, esplosivo, mentre durate piu' lunghe danno un'impressione molto piu' sfumata. Quando i grani vengono distribuiti in una larga regione frequenziale, la nuvola ha un carattere massiccio, mentre se la banda e' stretta, ne risulta un suono dotato di altezza propria. Densita' sparse di grani danno un effetto puntiglioso.

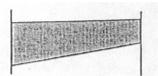


Figura 5.7: Esempio di maschera per la sintesi granulare asincrona

5.2.5 Sintesi sottrattiva

Mentre la sintesi additiva costruisce suoni complessi sommando insieme semplici suoni sinusoidali tempo varianti, la sintesi sottrattiva e' basata sull'idea complementare di passare un segnale a larga banda attraverso un filtro tempo variante per produrre la forma d'onda desiderata. La sintesi sottrattiva trae la sua origine nel campo analogico, dove si usava produrre segnali a partire da forme d'onda semplici, come onde quadre o a dente di sega e poi sagomare lo spettro mediante filtraggio eventualmente variabile. Nel campo numerico si ha il vantaggio di poter controllare in modo molto piu' preciso i parametri dei filtri. D'altra parte spesso conviene generare direttamente la forma d'onda voluta con altri metodi. Essa pertanto si e' piu' specializzata nell'uso con filtri piu' sofisticati o i cui parametri possano essere stimati a partire da suoni reali. Invece i filtri sono molto usati per produrre trasformazioni di suoni nel cosiddetto postprocessing. Mediante filtri cioe' si arricchisce un segnale sintetizzato o registrato di vari effetti, si variano le sue caratteristiche spettrali, pur mantenendo la

dinamica intrinseca del suono, si possono produrre effetti di riverberazione e spazializzazione e così via.

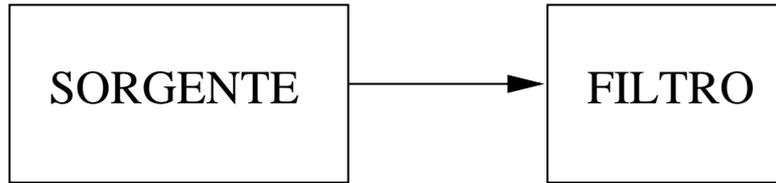


Figura 5.8: Sintesi sottrattiva

L'interpretazione fisica della sintesi sottrattiva consiste in una sorgente di segnale di eccitazione che viene inviata ad un sistema risonante (fig. 5.8). Questa descrizione si adatta in prima approssimazione a vari strumenti musicali tradizionali. Ad esempio le corde vibranti di un violino sono accoppiate attraverso il ponticello alla cassa risonante, che in questo caso si comporta come filtro tempo invariante. Anche il suono della voce può essere modellato come una sorgente di eccitazione, che può essere di tipo impulsivo data dalle vibrazioni delle corde vocali e rumorosa data dal flusso turbolento dell'aria in qualche costrizione del tratto vocale. Questa sorgente viene trasmessa attraverso il tratto vocale, la cavità orale, la cavità nasale, l'apertura delle labbra che filtrano e modificano spettralmente la sorgente, in modo approssimativamente lineare. Va osservato però che, nel caso della voce si può considerare la sorgente indipendente dal tratto vocale, mentre in molti strumenti, come il violino, per ottenere un modello efficace non si può trascurare l'influenza della risonanza sulla sorgente. La sintesi sottrattiva si applica bene nel primo caso, mentre nel secondo è bene ricorrere alle tecniche della sintesi per modelli fisici.

Nella sintesi sottrattiva si ha un blocco di generazione del segnale seguito in connessione senza feedback da uno o più blocchi, eventualmente in cascata o parallelo che lo trasformano. Questa tecnica di sintesi consiste nel considerare la trasformazione operata da un filtro su una sorgente di segnale, spettralmente ricca, come può essere quella effettuata dalla cassa di risonanza di un violino sulle vibrazioni delle corde.

I filtri lineari sono descritti dall'equazione alle differenze

$$y(n) = \sum_i b_i x(n-i) - \sum_k a_k y(n-k) \quad (5.4)$$

dove a_k e b_i sono i coefficienti del filtro e $x(n)$ e $y(n)$ sono rispettivamente i segnali di ingresso e uscita. Il filtro è caratterizzato da una risposta in frequenza definita da

$$H(f) = Y(f)/X(f)$$

dove $X(f)$ e $Y(f)$ sono gli spettri del segnale di ingresso e uscita.

A seconda dell'andamento della risposta in frequenza, si può variare l'andamento globale dello spettro del segnale in ingresso, ad esempio estraendo una piccola porzione del suo spettro. Se il filtro è statico, cioè se i parametri del filtro non variano, resta costante anche il suo effetto. Se invece i parametri sono tempo varianti, cambia anche la risposta in frequenza del filtro. In uscita si avrà una combinazione delle variazioni spettrali nel tempo del segnale in ingresso e di quelle del filtro. I parametri del filtro sono quindi scelti in base alla risposta in frequenza voluta e alla dinamica timbrica desiderata. Se viene usata per la sintesi è bene che il segnale di ingresso non sia di frequenza fissa, ma abbia ad esempio un po' di tremolo. Solo in questo modo infatti viene percepita

la forma dell'involuppo spettrale a causa delle variazioni d'ampiezza delle varie parziali che seguono l'involuppo spettrale.

La scomposizione effettuata offre la possibilità di controllare separatamente le caratteristiche della sorgente da quelle del filtro dando quindi una maggiore flessibilità nel controllo parametrico e una migliore interpretazione dei parametri di controllo.

Esaminiamo alcune tipiche applicazioni dei filtri nel campo audio.

Il filtro passa-basso (LP) con risonanza è usato spesso per simulare l'effetto di strutture risonanti; il filtro passa-alto (HP) invece per rimuovere componenti a bassa frequenza indesiderate; il filtro passa-banda (BP) può produrre effetti come imitazione di una linea telefonica, o la sordina in uno strumento musicale; il filtro elimina-banda (BR) può dividere lo spettro udibile in due bande separate che sembrino incorrelate. Il filtro risonante può essere usato per introdurre risonanze artificiali ad un suono; mentre il filtro notch (che elimina tutte le frequenze in una stretta banda attorno alla frequenza di risonanza) serve per eliminare disturbi quasi sinusoidali come ad esempio i 50 Hz dovuti all'alimentazione dei dispositivi elettronici. Un insieme di filtri notch usati in combinazione sul segnale di ingresso, può produrre l'effetto di phasing.

Va segnalato che il filtraggio può cambiare molto l'intensità del suono filtrato. Infatti il filtro può produrre l'effetto desiderato, ma il risultato non può poi essere usato perché diventato troppo debole o forte. Il metodo per compensare queste variazioni si chiama normalizzazione. In genere i metodi di normalizzazione impiegano norme del tipo L_1 , L_2 e L_∞ sul modulo della risposta in frequenza del filtro. La norma L_1 è usata quando il filtro non deve essere sovraccaricato in nessuna circostanza. Spesso però questo significa attenuare troppo il segnale. La norma L_2 (normalizzazione del valore efficace) è usata per normalizzare l'intensità del segnale. Questo metodo è accurato per segnali a larga banda e adatto in molte applicazioni musicali. La norma L_∞ normalizza la risposta in frequenza rispetto al suo massimo ed è efficace quando il segnale da filtrare è sinusoidale o periodico.

Un banco di filtri consiste in un gruppo di filtri che sono alimentati con lo stesso segnale. Ciascun filtro è tipicamente un filtro passa-banda stretto impostato ad una propria frequenza centrale. Spesso i segnali filtrati vengono poi sommati per produrre il suono in uscita. Quando si può controllare il livello di ciascun filtro il banco di filtri viene chiamato anche equalizzatore in quanto si può usare per compensare una risposta in frequenza non piatta del sistema di trasmissione o riproduzione.

Se si può controllare frequenza, banda e livello di ciascun filtro, si ha un sintetizzatore a formanti parallelo. Se le risposte dei singoli filtri non sono troppo sovrapposte, si riesce a controllare separatamente l'andamento dei singoli formanti. Questo può essere usato nella sintesi della voce, dove le transizioni tra i formanti devono essere accurate.

La tecnica vista si presta bene a sintetizzare sia gli involuppi spettrali poco variabili nel tempo, come le risposte acustiche ambientali, l'effetto delle casse armoniche, le spazializzazioni, sia gli involuppi spettrali rapidamente variabili, come gli effetti di sordina, la voce parlata e cantata e i suoni caratterizzati da grande dinamica timbrica. Si osservi che il modello non è limitato da assunti sulla periodicità del segnale sorgente, ma anzi può utilmente essere impiegato per la simulazione di segnali non intonati, come le percussioni. Per questi ultimi sono normalmente impiegate sorgenti di segnali rumorosi, caratterizzati da spettri continui. In quest'ultimo caso il modello sorgente di rumore bianco - filtro diventa un valido mezzo per descrivere i processi stocastici; esso infatti permette la caratterizzazione dell'involuppo spettrale, eventualmente considerato tempo-variante, che è il parametro percettivamente più significativo.

Se si possono fare ipotesi semplificative sull'ingresso, è possibile stimare sia i parametri della sorgente che del filtro a partire da un suono dato. La procedura più nota è il metodo di predizione lineare (LPC) che usa una sorgente composta da treno di impulsi o da rumore bianco ed è usata per la sintesi della voce. Questo metodo verrà presentato più estesamente nel paragrafo 5.2.6. Analizzando

una sequenza di segmenti di suono si ottengono parametri tempo varianti che possono essere usati nella sintesi. Il vantaggio di avere un modello parametrico e' che si puo' dare un'interpretazione fisica o spettrale a questi parametri e quindi avere un criterio di riferimento per la loro modificazione, sintetizzando quindi varianti del suono. Per esempio la stima dei parametri LPC della voce fornisce un filtro tempo variante che contiene l'andamento nel tempo dell'involuppo spettrale e quindi delle formanti. Questi sono parametri particolarmente importanti per la percezione della voce. Per cui essi possono essere modificati in senso spettrale cambiando il carattere della voce, o in senso temporale, facendo quindi una compressione o espansione della scala temporale, oppure si puo' cambiare il pitch della voce senza cambiare il suo involuppo e quindi mantenendo le caratteristiche della voce originaria. Una possibilita' usata spesso dai musicisti consiste nell'usare il filtro, con parametri stimati sul una voce parlata, applicando all'ingresso suoni d'altro tipo ricchi spettralmente. Vengono cosi' combinate le caratteristiche tempo-frequenza dei due suoni ottenendo, ad esempio, un'orchestra che canta. Questa tecnica viene chiamata sintesi incrociata.

Dal punto di vista implementativo va detto che l'operazione di filtraggio lineare puo' essere realizzata con diverse strutture che realizzano l'equazione alle differenze sopra vista, oppure come convoluzione con una risposta all'impulso, che di fatto descrive un filtro, o la risposta all'impulso di un ambiente. Una maniera alternativa consiste nel fare il filtraggio in frequenza, facendo il prodotto dello spettro del segnale, suddiviso in blocchi, con la risposta in frequenza del filtro, e antitrasformando il risultato.

La sintesi sottrattiva fa riferimento ad una interpretazione in frequenza. Le varie tecniche di implementazione dei filtri offrono diverse possibilita' di controllo parametrico, che vanno scelte in base alle applicazioni. Ad esempio per la sintesi di suoni vocalici e' utile poter controllare la frequenza e la banda dei formanti, dove si concentra l'energia e che caratterizza specialmente l'identita' del suono. Inoltre spesso e' utile combinare sorgenti periodiche a sorgenti stocastiche.

Lo stesso tipo di trasformazioni lineari, descritte da (5.4), puo' essere usato per ottenere effetti di riverberazione e periodizzazione. In questo caso il filtro e' caratterizzato da ritardi consistenti, che vengono meglio interpretati nel tempo, come echi, riverberi o come ripetizioni periodiche del segnale in ingresso. Metodi che impiegano filtri caratterizzati da delay lunghi sono presentati nel capitolo relativo alla spazializzazione.

Esempio di sintesi sottrattiva in Matlab

La sintesi sottrattiva crea suoni selezionando porzioni dello spettro di un segnale sorgente. E' dunque necessario disporre di segnali sorgente adeguati (cioe' a largo spettro) e saper progettare accuratamente i filtri numerici per la modellazione dello spettro.

In linea di principio, qualsiasi segnale puo' essere usato come sorgente di un algoritmo di sintesi sottrattiva. Esistono tuttavia due generatori di segnale notevoli che, per la ricchezza spettrale dei segnali prodotti, sono considerati particolarmente adatti a questo scopo: il *generatore di rumore* e il *generatore di impulsi*. Il primo produce un segnale non periodico e a spettro continuo su tutta la banda di frequenze. Il secondo produce un segnale periodico con energia distribuita uniformemente su tutta la banda di frequenze disponibile. Ogni linguaggio dispone di una funzione per la generazione di numeri casuali (in MATLAB e' presente la funzione `rand()`) che puo' essere utilizzata per generare rumore bianco. Non sempre e' previsto, invece, un generatore di impulsi. Il modo piu' semplice di costruire un treno di impulsi e' quello di usare la sintesi additiva per sommare tutte le componenti armoniche, multiple di una fondamentale che definisce il pitch, comprese fra 0 e $\frac{F_s}{2}$. In fig. 5.9 e' illustrato il segnale nel tempo e in frequenza.

La teoria del progetto dei filtri numerici e' argomento vasto e non e' affrontato in questa sede. A

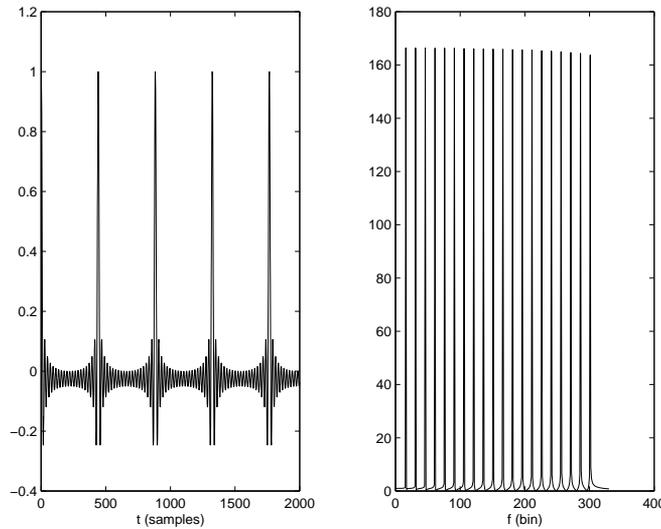


Figura 5.9: Generatore di impulsi

titolo di esempio e' invece tratto il caso notevole del progetto di celle IIR del secondo ordine. Come esercizio e' possibile realizzare le funzioni per le restanti celle FIR e IIR di primo e secondo ordine.

Le relazioni che descrivono la cella del secondo ordine sono le seguenti

$$\begin{aligned}
 y(n) &= b_1x(n) - a_2y(n-1) - a_3y(n-2) \\
 H(z) &= \frac{b_1}{1 + a_2z^{-1} + a_3z^{-2}} \\
 &= \frac{b_1}{1 - 2r\cos(\omega_c)z^{-1} + r^2z^{-2}}
 \end{aligned}$$

dati B larghezza di banda, f_c frequenza centrale della risonanza e $\omega_c = 2\pi f_c/F_s$, i coefficienti del filtro si legano con buona approssimazione a questi parametri mediante le relazioni

$$\begin{aligned}
 r &= e^{\frac{-\pi B}{F_s}} \\
 a_2 &= -2r\cos\left(\frac{2\pi f_c}{F_s}\right) \\
 a_3 &= r^2 \\
 b_1 &= (1-r)\sqrt{1 - 2r\cos(2\omega_c) + r^2}
 \end{aligned}$$

dove b_1 , fattore di normalizzazione del guadagno, e' calcolato ponendo $|H(\omega_c)| = 1$.

Per realizzare filtri lineari tempo-invarianti il MATLAB mette a disposizione la funzione `filter(b, a, in)`, in cui i parametri b e a sono i coefficienti del filtro secondo l'equazione alle differenze standard $a_1y(n) = b_1x(n) + b_2x(n-1) + \dots + b_{n_b+1}x(n-n_b) - a_2y(n-1) - a_3y(n-2) - \dots - a_{n_a+1}y(n-n_a)$. Con questa funzione e' dunque semplice sperimentare l'effetto del filtraggio a parametri costanti. Ad esempio, si supponga di voler realizzare un sintetizzatore a formanti parallelo con tre formanti centrati alle frequenze caratteristiche che determinano il tipo di vocale per la voce. Le righe seguenti mostrano come usare tre celle IIR in parallelo in uno schema ad eccitazione impulsiva e filtri in parallelo (la funzione `baIIR2([f1, f2, ..., fn], [B1, B2, ..., Bn])` restituisce due

matrici con i coefficienti delle n celle del secondo ordine descritte dai vettori di ingresso, mentre la funzione `buzz(t0, a, f)` realizza il generatore di impulsi).

```
% genera segnali di controllo
f=envgen([0,.2,1.8,2],[200,250,250,200],'linear'); % curva di pitch
% vibrato:
f=f+max(f)*0.05*sin(2*pi*5*(SpF/Fs)*[0:length(f)-1]).*hanning(length(f))';
amp=envgen([0,.2,1,1.8,2],[0,1,.8,1,0],'linear');

%genera sorgente impulsiva
s=buzz(0,amp,f);

%involuppo spettrale /i/
[b,a]=baIIR2([300 2400 3000],[200 200 500]);
si=filter(b(1,:),a(1,:),s)+
    filter(b(2,:),a(2,:),s)+
    filter(b(3,:),a(3,:),s);

%involuppo spettrale /a/
[b,a]=baIIR2([700 1200 2500],[200 300 500]);
sa=filter(b(1,:),a(1,:),s)+
    filter(b(2,:),a(2,:),s)+
    filter(b(3,:),a(3,:),s);

%involuppo spettrale /e/
[b,a]=baIIR2([570 1950 3000],[100 100 800]);
se=filter(b(1,:),a(1,:),s)+
    filter(b(2,:),a(2,:),s)+
    filter(b(3,:),a(3,:),s);
```

La funzione `baIIR2` calcola i coefficienti del filtro data la frequenza centrale e banda

```
function [b,a]=baIIR2(fc,B);
%funzione per la valutazione di coefficienti della cella del II ordine

global Fs;
global Fc;
nfilters=length(fc);

r=exp(-(pi.*B)/Fs)
a2=-(2*r.*cos(2*pi*fc/Fs))'
a3=r'.^2
a1=ones(nfilters,1)

%normalizzazione
b1=(1-r).*sqrt(1-2.*r.*cos(2*2*pi.*fc/Fs)+r.*r);
b1=b1';

a=[a1 a2 a3];
b=[b1 zeros(nfilters,1) zeros(nfilters,1)];
```

In figura 5.10 sono messi a confronto lo spettro del segnale originario con gli spettri del segnale con involuppo spettrale dato dalla sezione di filtri.

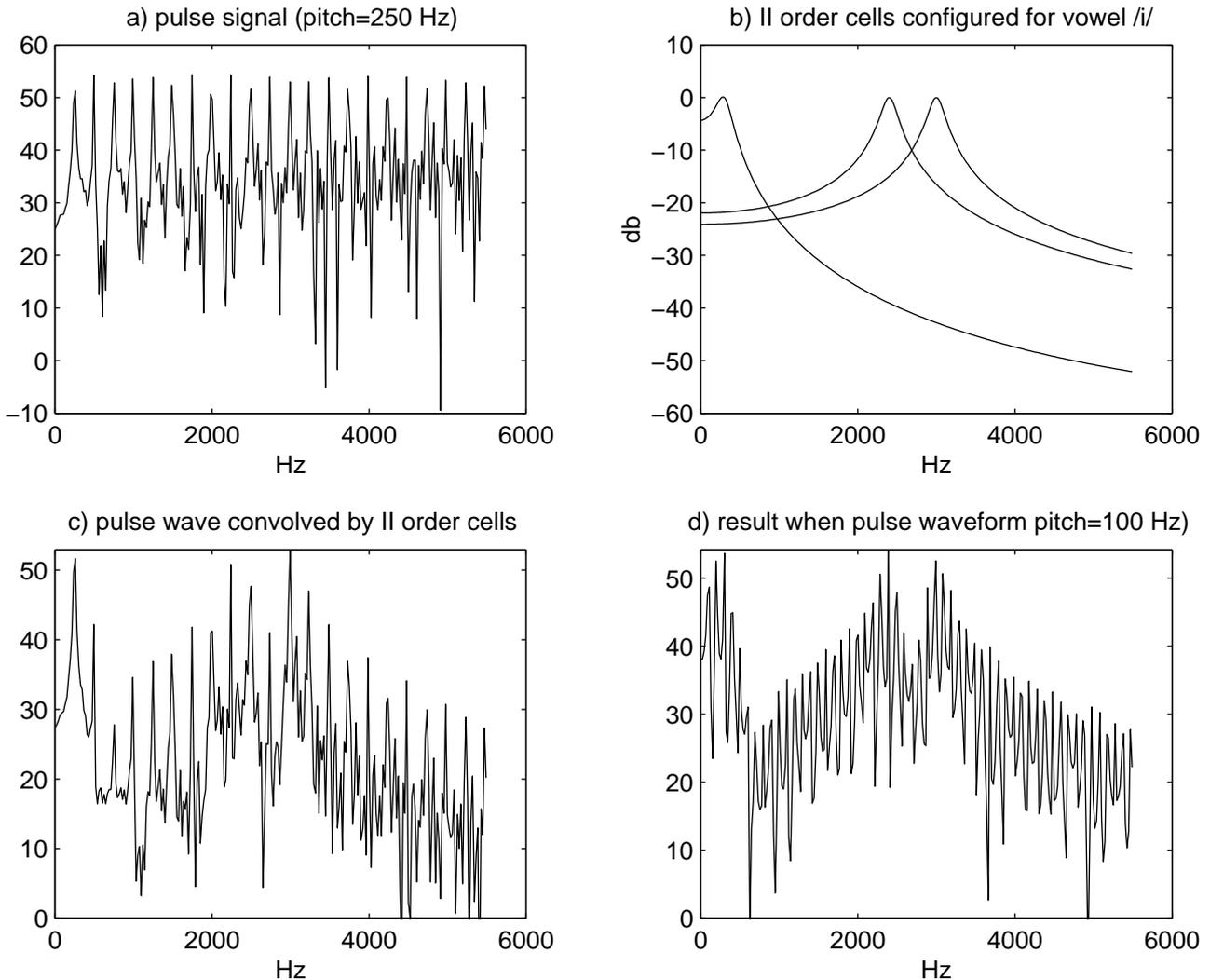


Figura 5.10: Involuppo formantico con banco di filtri in parallelo

Il banco di filtri puo' naturalmente essere applicato anche al suono campionato di uno strumento acustico. Questo tipo di operazione e' adatta, per esempio, a produrre effetti di variazione timbrica a partire dal timbro originario di una nota dello strumento da riprodurre. Un caso tipico e' la riproduzione dell'effetto di una sordina su uno strumento a fiato. Di seguito e' riportato un semplice esempio di manipolazione timbrica di un file audio.

```
% carica file audio
s=wavread('sample.wav');

[b,a]=baIIR2([300 2400 3000],[200 200 200]);
si=filter(b(1,:),a(1,:),s)+filter(b(2,:),a(2,:),s)
    +filter(b(3,:),a(3,:),s);
```

Nei casi in cui si vogliono ottenere cambiamenti di timbro continui nel tempo, e' necessario cambiare i parametri del filtro nel tempo. Questa e' una operazione non banale che da' occasione di accennare ad alcuni problemi fondamentali dei filtri tempo-varianti: il verificarsi di discontinuita' e

transitori spuri nel suono di uscita a fronte di variazione a scalino dei parametri; l'interpolazione lineare dei parametri non sempre basta a risolvere i problemi ed e' necessario scegliere accuratamente le strutture realizzative dei filtri che presentano maggior robustezza alle variazioni parametriche. Per sperimentare l'uso dei filtri tempo-varianti si fornisce una versione con controllo al frame rate della cella IIR del secondo ordine (`IIRcell(fc,B,in)`).

```
function out = IIRcell(in,fc,B);
% cella MA 1 ordine
% f0 al control rate
% sin audio rate;
(...)

[b,a]=baIIR(fc,B);

initstate=zeros(1,2);
out=zeros(1,nframes*SpF);
for (i=1:nframes)
    framein=in(((i-1)*SpF+1):i*SpF);
    [out(((i-1)*SpF+1):i*SpF),endstate]=
        =filter(b(i,:),a(i,:),framein,initstate);
    initsate=endstate; %aggiornamento dello stato del filtro
end
```

5.2.6 Sintesi della voce per predizione lineare

L'apparato di fonazione

La voce umana e' prodotta dal flusso di aria attraverso l'apparato di fonazione. Esso e' composto da tre cavita' principali: la cavita' nasale, la cavita' orale e la cavita' faringale, schematizzate in fig. 5.11. La cavita' nasale e' principalmente ossea e quindi la sua forma e' fissa. Essa puo' essere isolata dal resto dell'apparato vocale se si solleva il velo palatino, o palato molle. Così facendo si chiude il diaframma rinovelare che mette in comunicazione la cavita' nasale con quella orale e faringale. Quando l'apparato vocale e' in posizione di riposo, il velo pende giu' e il diaframma e' quindi aperto. Durante la produzione della maggior parte dei suoni linguistici il velo e' sollevato e il diaframma e' chiuso, ma nel caso di suoni nasali o nasalizzati esso rimane aperto, in modo che l'aria sfugge attraverso la cavita' nasale, conferendo al suono una caratteristica colorazione nasale.

La sommita' della cavita' orale e' formata dalla struttura ossea del palato e dal palato molle. La conformazione della cavita' puo' essere modificata in modo considerevole dal movimento della mandibola, che puo' aprire o chiudere la bocca; dalle labbra, la cui disposizione puo' variare dall'estremo appiattimento all'estremo arrotondamento; dalla lingua che puo' assumere una quantita' di posizioni diverse.

La cavita' faringale si estende fino al fondo della gola. Essa puo' essere compressa ritraendo indietro la radice della lingua verso la parete della faringe. Nella sua parte inferiore essa termina con le corde vocali, una coppia di membrane carnose che l'aria attraversa provenendo dai polmoni. Lo spazio tra esse e' detto glottide. Durante la produzione di un suono, essa puo' essere completamente aperta, con le corde vocali in posizione di quiete, parzialmente chiusa con le corde vocali in vibrazione o completamente chiusa, isolando così la cavita' faringale dai polmoni.

La forma d'onda del segnale vocale e' quella di un'onda di pressione acustica originata da movimenti fisiologici dell'apparato di fonazione. L'aria e' spinta dai polmoni nella trachea e quindi forzata

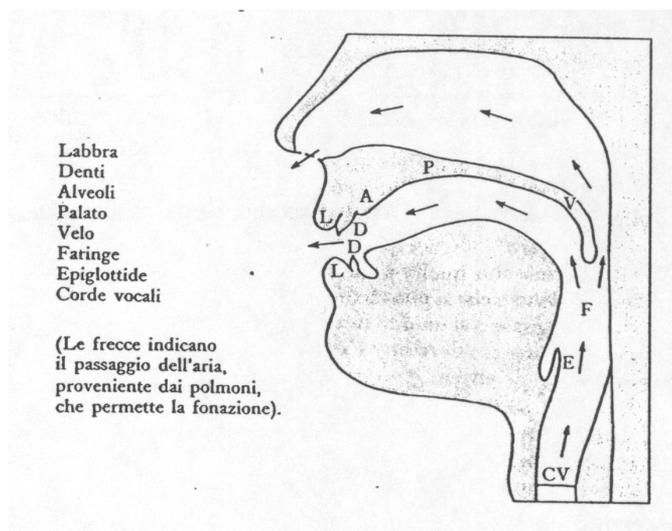


Figura 5.11: L'apparato di fonazione

attraverso le corde vocali. Durante la generazione di suoni *vocalizzati* (quasi periodici), come la vocale /a/, l'aria spinta dai polmoni causa la vibrazione delle corde vocali e quindi la modulazione del flusso d'aria ad una frequenza dipendente dalla pressione nella trachea e dalla disposizione (lunghezza, spessore, tensione) delle corde vocali. Più grande è la tensione delle corde vocali, più alta è la frequenza della voce. La velocità volumetrica (portata) del flusso di aria che attraversa la glottide definisce l'ingresso o eccitazione del tratto vocale.

I suoni non vocalizzati come /f/ sono generati tenendo volontariamente aperte le corde vocali, forzando l'aria attraverso la glottide e quindi usando l'articolazione per creare una costrizione lungo il tratto vocale (ad esempio posando i denti superiori sul labbro inferiore per il fonema /f/). Con contemporanea costrizione e vibrazione delle corde vocali si generano le fricative vocalizzate come la /z/ di rosa o la /dz/ di zanzara. I suoni esplosivi come /p/ sono generati aumentando la pressione dell'aria nella bocca e facendola quindi uscire improvvisamente.

Alla produzione di ogni fonema corrisponde una certa configurazione anatomica del tratto vocale, il quale agisce come risonatore meccanico allo scopo di modificare lo spettro dell'eccitazione glottale. Le frequenze di risonanza del tratto vocale vengono dette formanti. Ogni suono è caratterizzato dal valore assunto dalle formanti che tipicamente sono in numero di quattro nell'intervallo 0-4 kHz. Nel parlato continuo, la configurazione del tratto vocale varia nel tempo, per cui l'evoluzione temporale delle formanti costituisce un efficace metodo di rappresentazione del segnale vocale. Inoltre, considerando che la velocità di variazione degli organi articolatori è abbastanza lenta, è possibile schematizzare il processo come stazionario a tempo breve (dell'ordine di 10 - 50 msec).

Un modello di analisi e sintesi del parlato: la predizione lineare

Come esempio di sintesi sottrattiva presentiamo ora un algoritmo semplificato per la sintesi della voce. Un modello lineare di produzione della voce è stato sviluppato da Fant nel 1950. In esso la velocità volumetrica che esce dalla glottide è modellata come uscita di un filtro passa-basso $G(z)$ a due poli con frequenza di taglio stimata a circa 100Hz. L'ingresso $u(t)$ del filtro è un treno di impulsi con frequenza f_0 per i suoni vocalizzati e rumore bianco per i suoni non vocalizzati. Il tratto vocale è modellato con un filtro $V(z)$ a soli poli, consistente in una cascata di un numero ridotto di risuonatori

(filtri) del secondo ordine. Ogni risonanza è definita come un *formante* con una frequenza centrale f_i e una larghezza di banda B_i . Un modello più accurato dovrebbe comprendere un infinito numero di risonanze, il cui effetto principale alle frequenze più basse è di aumentare il livello spettrale. Quindi quando si deve rappresentare accuratamente solo il funzionamento del sistema solo alle frequenze medio-basse, (la parte più importante per la percezione del parlato), è necessario introdurre una correzione che rappresenti l'effetto dei poli più alti trascurati.

La forma d'onda della velocità volumetrica alle labbra è trasformata nella forma della pressione acustica fuori dalle labbra per mezzo di un modello $L(z)$ di radiazione delle labbra. In definitiva si ha

$$S(z) = U(z) \cdot [G(z) \cdot V(z) \cdot L(z)] \quad (5.5)$$

dove $S(z)$ è il segnale prodotto.

Il modello della glottide è della forma

$$G(z) = \frac{1}{[1 - \exp(-c/F_c)z^{-1}]^2}$$

e il modello di radiazione delle labbra è della forma

$$L(z) = 1 - z^{-1}$$

Il modello dell' i -esimo formante di frequenza f_i e banda B_i è dato da

$$V_i(z) = \frac{g_i}{1 - 2r \cos(\theta_i)z^{-1} + r^2}$$

con $\theta_i = 2\pi f_i/F_c$ e $r = \exp(-\pi B_i/F_c)$ e g_i coefficiente di guadagno del filtro. Il modello cascata del tratto vocale è dato quindi da;

$$V(z) = \prod_{i=1}^K V_i(z)$$

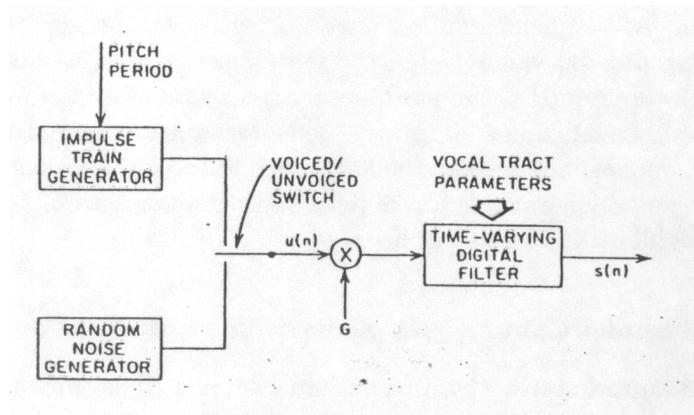


Figura 5.12: Diagramma a blocchi del modello semplificato di produzione del parlato

Il modello descritto prevede come ingresso un treno di impulsi periodici o un rumore e come parametri le frequenze e le larghezze di banda dei formanti (fig. 5.12). Nella sintesi della voce questi parametri vengono aggiornati ogni 5-10 msec o all'inizio di ogni periodo di pitch (sintesi sincrona con il periodo). Si può osservare che il termine al numeratore $L(z) = 1 - z^{-1}$ quasi coincide con

un polo di $G(z)$ che vale $1 - \exp(-c/F_c)z^{-1}$, in quanto $c/F_c \ll 1$. Si puo' quindi approssimare la composizione degli effetti spettrali della radiazione, tratto vocale ed eccitazione glottale con un filtro senza zeri e $p = 2K + 1$ poli. Esso e' pertanto rappresentabile con

$$S(z) = H(z) \cdot U(z) = \frac{g}{1 - \sum_{k=1}^p a_k z^{-k}} U(z) \quad (5.6)$$

Il filtro $H(z)$ rappresenta le caratteristiche complessive ingresso uscita del modello. In definitiva i parametri che definiscono il modello di produzione del parlato qui visto sono i coefficienti di $A(z)$ e i parametri dell'ingresso vocalizzato o non, pitch e guadagno g . Per stimarli a partire da un segnale vocale si dovra' pertanto prima decidere se usare un generatore casuale (per suoni non vocalizzati) o periodico, poi si stimera' la frequenza della fondamentale e il guadagno. Infine si stimano i coefficienti a_k tramite algoritmi predittivi, che possono essere ricondotti essenzialmente a due classi: metodo della covarianza e metodo dell'autocorrelazione. Il metodo dell'autocorrelazione e' quello attualmente piu' usato per l'esistenza di algoritmi piu' robusti e piu' efficienti.

Il modello semplificato a soli poli e' una rappresentazione naturale per i suoni non nasali, mentre invece per i suoni nasali e fricativi si dovrebbe tener conto anche degli zeri. D'altra parte se l'ordine p e' sufficientemente alto, anche il modello a soli poli produce una buona rappresentazione per quasi tutti i suoni del parlato. Il grande vantaggio e' che i parametri possono essere stimati in modo semplice.

Dall'equazione 5.6 risulta che i campioni $s(n)$ sintetizzati sono legati all'ingresso dall'equazione

$$s(n) = \sum_{k=1}^p a_k s(n-k) + gu(n) \quad (5.7)$$

L'equazione 5.6 e' detta modello di sintesi, in quanto applicando a questo filtro l'ingresso si ottiene il suono vocale in uscita.

Per la stima dei parametri si considera sconosciuto l'ingresso $u(n)$. Si consideri di stimare l'uscita del sistema approssimativamente da una somma pesata dei campioni precedenti. Si ha quindi un predittore lineare di ordine p , con coefficienti \tilde{a}_k , definito dalla relazione

$$\tilde{s}(n) = \sum_{k=1}^p \tilde{a}_k s(n-k) \quad (5.8)$$

Viene quindi definito l'errore di predizione $e(n)$ (chiamato anche residuo) la differenza tra il valore attuale $s(n)$ e il valore predetto $\tilde{s}(n)$

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p \tilde{a}_k s(n-k) \quad (5.9)$$

Da questa relazione si vede che $e(n)$ e' l'uscita di un sistema con funzione di trasferimento

$$A(z) = 1 - \sum_{k=1}^p \tilde{a}_k z^{-k}$$

e con ingresso i campioni del segnale vocale $s(n)$. Confrontando le equazioni 5.7 e 5.9 si vede che se il segnale seguisse esattamente il modello di eq. 5.7 e se $\tilde{a}_k = a_k$, allora risulterebbe

$$e(n) = gu(n) \quad (5.10)$$

In questo caso il filtro $A(z)$ di predizione dell'errore risulta coincidere con il filtro inverso del sistema $H(z)$ e cioè $H(z) = g/A(z)$.

Si stimano i parametri del modello direttamente dai campioni del segnale vocale cercando di ottenere anche una buona stima delle proprietà spettrali del segnale ottenuto utilizzando il modello per la sintesi (eq. 5.7). A causa della natura tempo variante del segnale vocale si farà la stima su segmenti corti del segnale o a blocchi di campioni. L'approccio che viene seguito si basa sul metodo dei minimi quadrati, minimizzando quindi il quadrato dell'errore $e(n)$ di predizione su un segmento di suono

$$E = \sum_m e^2(m)$$

dove la somma è estesa ai campioni del segmento analizzato. I parametri risultanti sono assunti essere i parametri della funzione del sistema $H(z)$ nel modello di produzione del parlato. Ricordando la relazione 5.10, si stima quindi g confrontando l'energia del segnale errore con quello scelto come eccitazione mediante la relazione

$$g^2 = \frac{\sum_m e^2(m)}{\sum_m u^2(m)}$$

Per minimizzare E , si ottengono le cosiddette equazioni di Yule-Walker, che consentono di determinare i coefficienti del filtro. Si noti che la minimizzazione ai minimi quadrati di E tende a produrre un segnale di errore con modulo dello spettro piatto (rumore bianco); per cui il filtro $A(z)$ è chiamato anche *whitening filter*. Se il modello approssima bene il segnale vocalizzato, allora il residuo è composto da un treno di impulsi che si ripetono alla frequenza di vibrazione delle corde vocali. Pertanto gli errori massimi di predizione si verificheranno con frequenza uguale al pitch del segnale. Nel dominio del tempo quindi la maggior parte dell'energia si localizza in vicinanza di questi picchi.

È interessante notare che l'interpretazione in frequenza di E comporta che il metodo LPC stimato con l'autocorrelazione approssima meglio lo spettro nelle regioni di segnale ad alta energia, cioè vicino ai picchi dello spettro rispetto alle regioni a bassa energia (valli). Un esempio è riportato in fig. 5.13.

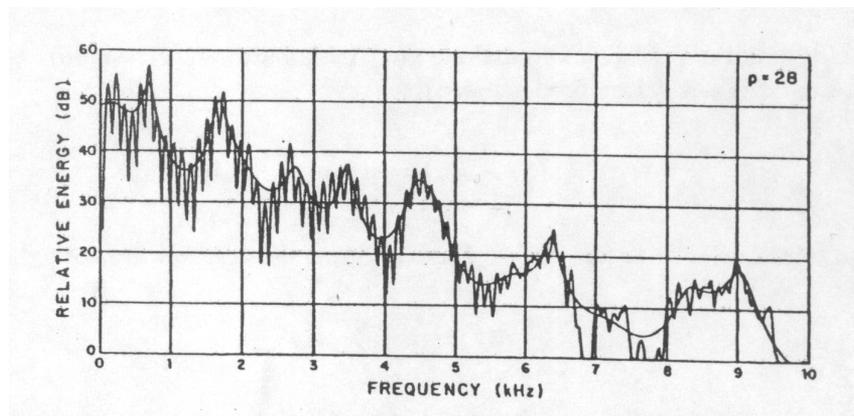


Figura 5.13: Spettro LPC con 28 poli confrontato con quello ottenuto da analisi mediante FFT

I parametri così stimati sono assunti essere i parametri del modello. Da essi si possono ricavare altri parametri percettualmente più significativi come la frequenza e banda dei formanti. Questo approccio ha il vantaggio di avere metodi di stima efficienti e che si sono rivelati anche fornire una rappresentazione accurata del segnale vocale.

Esempio di analisi e risintesi LPC in MATLAB

Per l'analisi LPC dei segnali il MATLAB mette a disposizione la funzione `lpc(s,N)`, dove s è il segnale di ingresso e N è l'ordine del filtro predittivo. Nell'esempio che segue, un frammento di segnale vocale viene analizzato per calcolare i coefficienti del filtro. Il filtro inverso è poi usato per ricavare l'eccitazione glottale e, infine, il segnale vocale di partenza è ricalcolato per filtraggio diretto dell'eccitazione glottale mediante il filtro predittivo.

```
%legge campione audio
svoce=wavread('voce.wav');

%selezione della finestra temporale da analizzare
s=svoce(8000:10000);

% calcolo dei coefficienti di predizione lineare
Nc=10; %numero coefficienti
[a,g]=lpc(s,Nc); %a -> coefficienti, g -> gain

freqz([g 0 0],[a]); %plot della risposta del filtro

% Generazione dell'eccitazione glottale
% mediante filtraggio inverso del segnale vocale
u=filter([a],[g 0 0],s);

% risintesi
snew=filter([g,0,0],[a],u);
```

La figura 5.14 illustra la risposta in frequenza del filtro $A(z)$, l'eccitazione glottale e la forma d'onda ottenuta mediante risintesi LPC.

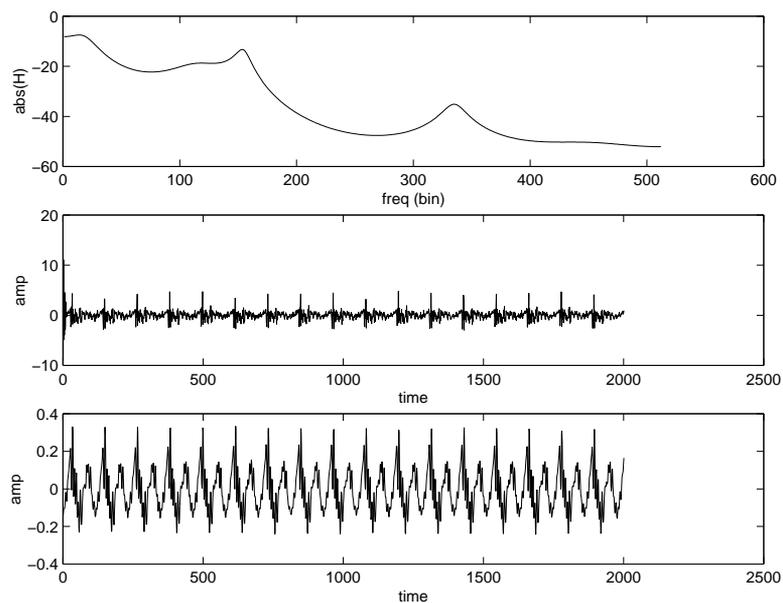


Figura 5.14: Analisi e risintesi della voce mediante LPC

5.3 Sintesi non lineari

Le trasformazioni viste sopra non possono cambiare le frequenze delle componenti in ingresso, in quanto sono trasformazioni lineari. Se si usano invece trasformazioni non lineari, le frequenze possono cambiare anche di molto. Ne consegue la possibilità di cambiare sostanzialmente la natura del suono in ingresso. Queste possibilità vengono anche usate nella sintesi del suono.

L'interpretazione della sintesi non lineare non è basata sull'acustica fisica, ma piuttosto deriva dalla teoria della modulazione nelle comunicazioni elettriche, applicata ai segnali musicali. Questi metodi sono stati molto usati nella musica elettronica analogica e sono poi stati anche sviluppati nel digitale. Pertanto la sintesi non lineare ne eredita parzialmente l'interpretazione analogica come usata nella musica elettronica e inoltre è diventata, specie con la modulazione di frequenza, una nuova metafora per i musicisti informatici.

Ci sono due effetti principali legati alla trasformazione non lineari: arricchimento dello spettro e traslazione dello spettro. Il primo effetto deriva dalla distorsione non lineare di un segnale e consente di controllare la brillantezza di un suono, mentre il secondo è dovuto alla sua moltiplicazione per una sinusoidale (portante) e sposta lo spettro attorno alla frequenza del segnale portante, alterando il rapporto armonico tra le righe del segnale modulante. La possibilità di traslare lo spettro è molto efficace nelle applicazioni musicali. A partire da semplici componenti, si possono creare suoni armonici e inarmonici e stabilire differenti relazioni armoniche tra le parziali.

5.3.1 Sintesi per modulazione di frequenza

I due metodi classici per l'arricchimento spettrale e per la traslazione dello spettro, ovvero la distorsione non lineare (vedi par. 5.3.4) e modulazione ad anello (vedi par. 5.3.3), hanno perso progressivamente di interesse in favore del metodo della modulazione in frequenza, il quale unisce i due effetti ed evita alcuni difetti di questi metodi. Per questa ragione, la sintesi per modulazione di frequenza, proposta da J. Chowning nel 1973 e' il più usato fra i metodi non lineari. Questo metodo è diventato molto popolare da quando fu usato nei sintetizzatori Yamaha tipo DX7 ed è tuttora usato anche nelle schede audio. Questa tecnica non deriva da modelli del segnale o di produzione del suono, ma piuttosto dalle proprietà matematiche di una formula. In realtà con questo nome non si intende una sola tecnica, ma tutta una famiglia che è costituita dalla modulazione della fase o frequenza istantanea di una sinusoidale (portante) secondo l'andamento di un altro segnale (modulante), spesso sinusoidale. Anche se nella computer music si usa far riferimento a questa tecnica con il nome di modulazione di frequenza, in quanto le prime realizzazioni erano di questo tipo, conviene sviluppare la teoria facendo riferimento alla formulazione come modulazione di fase.

Il modulo di calcolo base è dato da:

$$s(t) = \sin(2\pi f_c n + \phi(t)) \quad (5.11)$$

dove $\phi(t)$ è il segnale di ingresso (modulante) e f_c è la frequenza della portante (*carrier*). Quando $f_c = 0$ si vede che risulta una distorsione non lineare del segnale modulante ed è pertanto da aspettarsi un arricchimento spettrale. Se invece $f_c \neq 0$ si può far vedere con le formule di prostaferesi che l'espressione scritta equivale alla modulazione ad anello di una sinusoidale per il segnale distorto (traslazione dello spettro di un valore f_c).

Modulante semplice

Se la modulante e' una sinusoidale di ampiezza I (indice di modulazione) e frequenza f_m data quindi da

$$\phi(t) = I \sin(2\pi f_m t)$$

la modulazione di frequenza semplice da':

$$\begin{aligned} s(t) &= \sin[2\pi f_c t + I \sin(2\pi f_m t)] \\ &= \sum_{k=-\infty}^{\infty} J_k(I) \sin[2\pi(f_c + k f_m)t] \end{aligned} \quad (5.12)$$

dove $J_k(I)$ e' la funzione di Bessel del primo tipo di ordine k . Dall'equazione 5.12 si vede che il segnale prodotto ha uno spettro a righe di frequenza $f_c \pm k f_m$ e di ampiezza data da $J_k(I)$. Pur essendo la sommatoria estesa ad un numero infinito di termini, solo pochi di essi, attorno a $k = 0$ sono significativi. Infatti solo le funzioni di Bessel di ordine basso sono significative per valori piccoli dell'indice di modulazione. Quando l'indice I cresce, aumenta in corrispondenza anche il numero di funzioni significative. Il numero M di frequenze laterali di ampiezza maggiore di un centesimo e' dato da $M = I + 2.4 \cdot I^{0.27}$. In pratica si puo' considerare $M = 1.5 * I$. In questo modo si controlla la larghezza di banda attorno a f_c . Ne risulta un effetto tipo filtro dinamico, analogo a quello che i musicisti sperimentano nell'impiego della sintesi sottrattiva. Inoltre l'ampiezza di ogni funzione varia in modo oscillante al variare dell'indice. Questo fatto produce una caratteristica ondulazione delle ampiezze delle parziali componenti quando l'indice varia in modo continuo e consente di ottenere facilmente spettri dinamici.

Nel caso base il segnale modulante e' sinusoidale e ne risulta uno spettro a righe di frequenza $|f_c \pm k f_m|$ caratterizzato quindi dal rapporto tra le frequenze. Si ha quindi la possibilita' di controllare il grado di inarmonicita' del segnale mediante il rapporto f_c/f_m , come verra' discusso nel paragrafo 5.3.2. Vale inoltre la proprieta' che l'ampiezza massima e la l'energia non cambiano al variare dell'indice I . Pertanto vengono evitati i problemi di normalizzazione di ampiezza che ci sono nella sintesi per distorsione non lineare.

E' interessante ora vedere l'equivalente formulazione di 5.12 come modulazione di frequenza. La frequenza istantanea di un segnale $s(t) = \sin[\psi(t)]$ e' data da $f(t) = [1/2\pi][d\psi(t)/dt]$. Pertanto la frequenza istantanea $f_i(t)$ del segnale di eq. 5.12 vale

$$f_i(t) = f_c + I f_m \cos(2\pi f_m t) \quad (5.13)$$

Essa varia quindi attorno a f_c con una deviazione massima $d = I \cdot f_m$. In figura 5.15 e' riportato il caso di modulazione con portante semplice realizzato mediante oscillatore controllato in frequenza. Si osservi infine che un cambio della differenza di fase tra portante e modulante produce solo un cambiamento delle fasi reciproche delle parziali generate. Questo normalmente non e' percettualmente significativo. Solo nel caso in cui alcune parziali coincidano in frequenza, bisogna tenere conto della loro relazione di fase per calcolare l'ampiezza risultante.



Figura 5.15: Modulazione di frequenza con modulante semplice

Esempio in Matlab E' possibile realizzare gli algoritmi per la modulazione di frequenza utilizzando la formula del modulo di calcolo base con modulante sinusoidale (equazione 5.12)¹:

```

Fs=22050; % frequenza di campionamento

fc=700; %portante: 700 Hz
fm=100; %modulante: 100 Hz
I=2; %indice di modulazione
t=0:(1/Fs):3; %asse temporale: 3 sec

s=sin(2*pi*fc*t+I*sin(2*pi*fm*t));

```

La figura 5.16 illustra lo spettro del segnale generato per tre valori diversi dell'indice di modulazione. Si puo' verificare come al crescere dell'indice di modulazione l'energia della portante si distribuisce progressivamente sulle bande laterali seguendo l'andamento previsto dalle funzioni di Bessel.

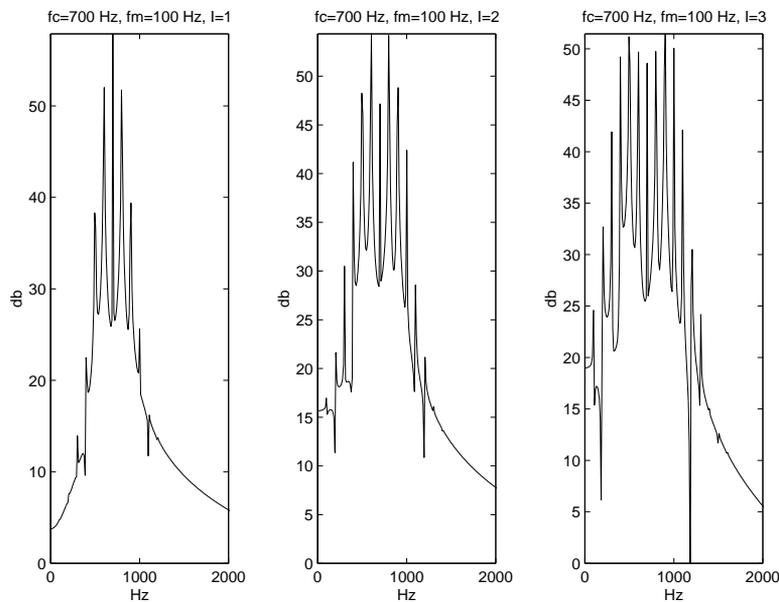


Figura 5.16: Spettro relativo a modulazione di frequenza con portante a 700 Hz, modulante sinusoidale a 100 Hz e indice di modulazione I crescente da 1 a 3

Portante composta

Consideriamo ora una portante periodica ma non sinusoidale.

$$s(t) = \sum_{l=0}^L A_l \sin[2\pi l f_c t + \phi_l(t)]$$

¹Si noti che la formula $s(n) = \sin(2\pi f_c n + \phi(n))$ corrisponde in realta' ad una modulazione di fase in cui la fase istantanea e' $\phi_i(n) = \phi_c + \phi(n)$. Nel caso di modulante sinusoidale ($\phi(n) = I \sin(2\pi f_m n)$), e' possibile verificare l'equivalenza tra modulazione di fase e modulazione di frequenza se in $2\pi f_i(n) = 2\pi(f_c + f(n))$ si pone $f(n) = I f_m \cos(2\pi f_m t)$. L'implementazione della modulazione in questa ultima forma richiede pero' attenzione ai problemi di continuita' di fase quando la frequenza e' imposta ad ogni istante

Se essa viene modulata, e' come se ciascuna sua armonica fosse modulata dalla stessa modulante. Se la modulante e' sinusoidale, nello spettro attorno ad ogni armonica della portante saranno presenti righe di ampiezza proporzionale all'armonica. Ne risulta uno spettro di righe a frequenza $|lf_c \pm kf_m|$ e di ampiezza $A_l J_k(I)$ con $|l| \leq L$ e $|k| \leq M$, essendo L il numero di armoniche significative.

$$s(t) = \sum_{l=1}^L \sum_{k=-M}^M A_l J_k(I) \sin[2\pi(lf_c + kf_m)t]$$

In generale ci possono essere varie portanti indipendenti modulate dalla stessa modulante o da differenti modulanti (fig. 5.17). Ne risulta una specie di sintesi additiva in cui invece che addendi sinusoidali, si hanno addendi piu' complessi

$$s(t) = \sum_{l=0}^L A_l \sin[2\pi f_c l n + \phi_l(t)]$$

Per esempio con portanti di frequenza multipla della frequenza della modulante f_m si possono creare suoni armonici complessi di frequenza fondamentale $f_0 = f_m$ controllando le varie regioni dello spettro in modo indipendente. La frequenza di ciascuna portante determina la regione che viene influenzata e in un certo senso la posizione di un formante.

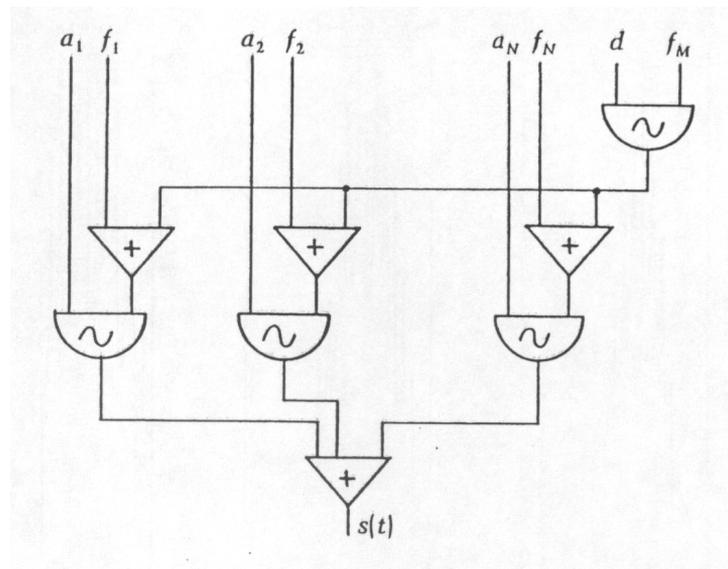


Figura 5.17: Modulazione di frequenza con N portanti modulate dalla stessa modulante

Esempio in Matlab L'esempio precedente puo' essere gradualmente arricchito considerando portanti e/o modulanti via via piu' complesse. L'uso di una portante composta, ad esempio, puo' servire per generare nello spettro dei formanti alle frequenze delle sue componenti. L'istruzione di generazione del segnale audio diventa

```
fc1=300; %portante 1: 300 Hz
fc2=1000; %portante 2: 1000 Hz
fc3=3000; %portante 3: 3000 Hz
```

```

fm=100; %modulante: 100 Hz
I1=1;
I2=2;
I3=3;
theta=sin(2*pi*fm*t);
s=sin(2*pi*fc1*t+I1*theta)
  +sin(2*pi*fc2*t+I2*theta)
  +sin(2*pi*fc3*t+I3*theta);

```

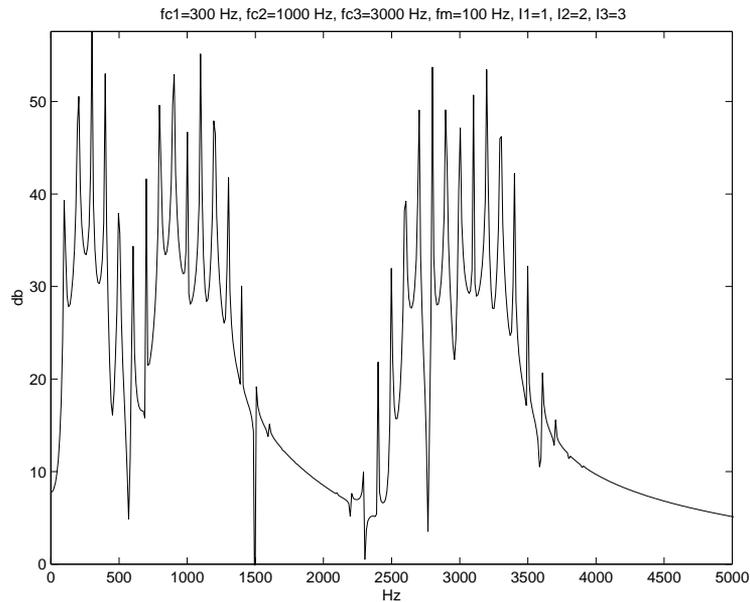


Figura 5.18: Spettro relativo a modulazione di frequenza con tre portanti e una modulante

Modulante composta

Esaminiamo ora il caso di modulante composta da due sinusoidi (fig. 5.19), ciascuna con il suo indice di modulazione

$$\phi(t) = I_1 \sin(2\pi f_1 t) + I_2 \sin(2\pi f_2 t)$$

Sostituendo in (5.11) risulta:

$$s(t) = \sin[2\pi f_c t + I_1 \sin(2\pi f_1 t) + I_2 \sin(2\pi f_2 t)]$$

Sviluppando la prima modulante si ha:

$$s(t) = \sum_k J_k(I_1) \sin[2\pi(f_c + k f_1)t + I_2 \sin(2\pi f_2 t)]$$

e poi la seconda modulante si arriva a:

$$s(t) = \sum_k \sum_n J_k(I_1) \cdot J_n(I_2) \sin[2\pi(f_c + k f_1 + n f_2)t]$$

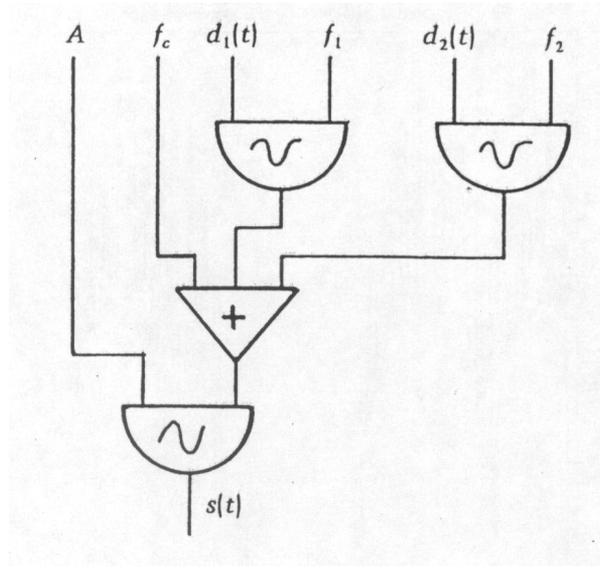


Figura 5.19: Modulazione di frequenza con due modulatori

Lo spettro risultante è molto più complicato di quello del caso di una modulante semplice. Sono presenti tutte le parziali a frequenza $|f_c \pm kf_1 \pm nf_2|$ e con ampiezza $J_k(I_1) \cdot J_n(I_2)$. Per interpretare l'effetto si consideri $f_1 > f_2$. Se fosse presente solo la modulante a frequenza f_1 , lo spettro risultante avrebbe un certo numero di componenti di ampiezza $J_k(I_1)$ e frequenza $f_c \pm kf_1$. Quando viene applicato anche la modulante a frequenza f_2 , queste componenti diventano a loro volta portanti con bande laterali prodotte da f_2 . Attorno a ciascuna delle componenti prodotte da f_1 si avranno cioè righe spaziate di f_2 . La banda risultante è approssimativamente uguale alla somma delle due bande.

Se le frequenze hanno rapporti semplici tra loro, lo spettro è del tipo $|f_c \pm kf_m|$ dove ora f_m è il massimo comun divisore tra f_1 e f_2 . Per esempio se $f_c = 700$ Hz, $f_1 = 300$ Hz e $f_2 = 200$ Hz, le componenti sono $700 \pm k100$ e la fondamentale 100 Hz. Pertanto scegliendo f_1 e f_2 multipli di f_m si ottengono suoni dello stesso tipo di quelli ottenuti con la modulazione semplice ma con una dinamica spettrale più ricca.

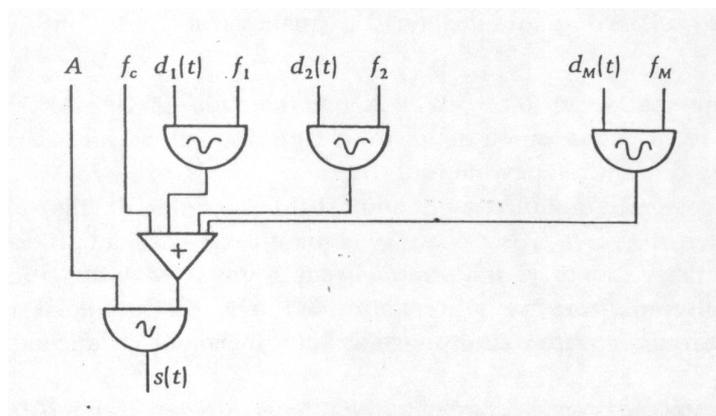


Figura 5.20: Modulazione di frequenza con N modulatori

In generale se il segnale modulante è composto da N sinusoidi (fig. 5.20) risulteranno componenti

di frequenza $|f_c \pm k_1 f_1 \pm \dots \pm k_N f_N|$ con ampiezze date dal prodotto di N funzioni di Bessel. Anche qui se i rapporti sono semplici risulta uno spettro del tipo $|f_c \pm k f_m|$ dove f_m e' il massimo comun divisore tra le frequenze modulanti. Se i rapporti non sono semplici le righe risultanti saranno sparse dando luogo a suoni inarmonici o anche rumorosi per alti valori degli indici.

Ad esempio Schottstaedt usa la doppia modulante per simulare il suono del piano, ponendo $f_1 \simeq f_c$ e $f_2 \simeq 4f_c$. In questo modo cerca di simulare la leggera inarmonicit  delle corde del piano. Inoltre fa diminuire gli indici di modulazione al crescere di f_c e quindi della fondamentale della nota. In questo modo le note basse sono piu' ricche di armoniche di quelle alte.

Esempio in Matlab La modulante puo' a sua volta essere composta da piu' componenti. Il caso di una portante sinusoidale e modulante composta si realizza ad esempio con

```
fc=700;
fm1=700;
fm2=2800;
I1=1;
I2=1;

s=sin(2*pi*fc*t+I1*sin(2*pi*fm1*t)+I2*sin(2*pi*fm2*t));
```

La figura 5.21 illustra due casi di modulazione con una portante sinusoidale a 700 Hz e una modulante composta da due sinusoidi. Si puo' osservare nella prima come i rapporti semplici delle frequenze di modulanti e portante determinino uno spettro del tipo $|f_c \mp k f_m|$ dove $f_m = 100\text{Hz}$ (massimo comun divisore tra $f_1 = 300$ e $f_2 = 200$) e la fondamentale e' a 100 Hz. Nella seconda, i rapporti di frequenza $f_1 = f_c$ e $f_2 = 4f_c$ sono scelti in modo che la fondamentale coincide con f_c e che le parziali superiori siano in rapporto armonico con la fondamentale, essendo proprio $f_1 = f_c$ il massimo comun divisore tra f_1 e f_2 (piano di Schottstaedt).

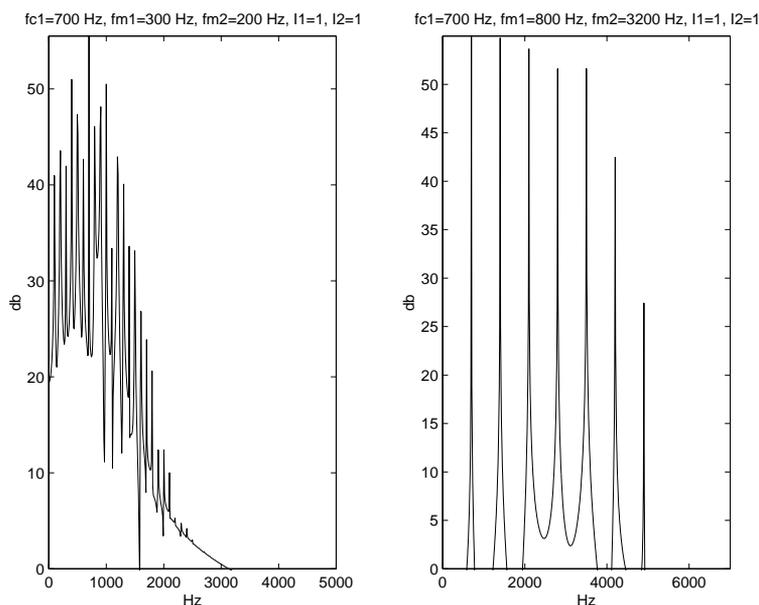


Figura 5.21: Due esempi di modulazione con portante semplice e modulante composta

I rapporti semplici dell'ultimo esempio visto determinano uno spettro esattamente armonico. E' possibile sperimentare l'effetto dell'inarmonicita' variando i valori di f_1 e f_2 in modo che siano solo approssimativamente pari a f_c e a $4f_c$ rispettivamente. La figura 5.22 mostra lo spettro risultante per scostamenti progressivi di f_1 e f_2 dai valori proporzionali a f_c .

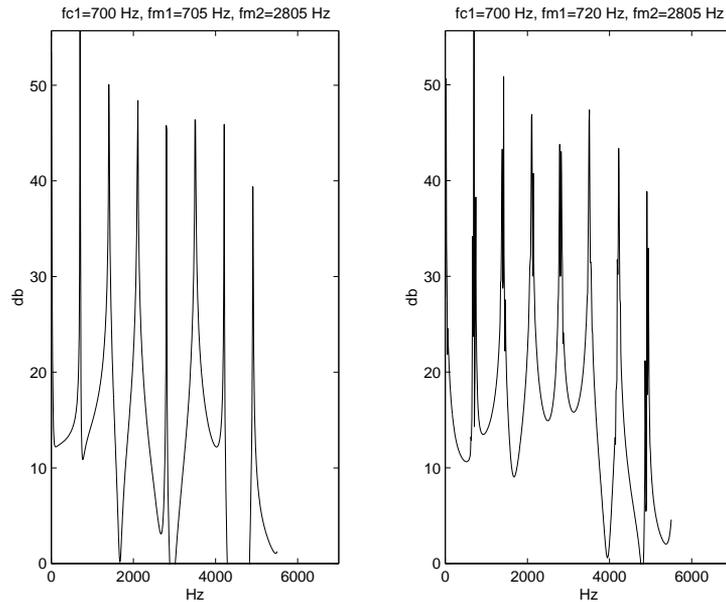


Figura 5.22: Suoni inarmonici dovuti a rapporti non semplici tra frequenze

Anche per gli algoritmi di modulazione di frequenza e' possibile pensare ad una interfaccia che renda semplice controllare la sintesi con involucri di ampiezza e frequenza al frame rate. Un oscillatore FM a portante e modulante composta, ad esempio, avrebbe interfaccia $FM_{oper}(t_0, a, [fc_1 fc_2 \dots fc_N], [fm_1 fm_2 \dots fm_M], [I_1 I_2 \dots I_M])$ in cui tutti i parametri di ingresso possono essere rappresentati con involucri temporali. La realizzazione di questo operatore e' lasciata come esercizio.

Modulanti in cascata

Consideriamo ora il caso di modulante sinusoidale a sua volta modulata da un'altra sinusoide (fig. 5.23)

$$\phi(t) = I_1 \sin(2\pi f_1 t + I_2 \sin(2\pi f_2 t))$$

Il segnale e' quindi definito da:

$$\begin{aligned} s(t) &= \sin[2\pi f_c t + I_1 \sin(2\pi f_1 t + I_2 \sin(2\pi f_2 t))] \\ &= \sum_k J_k(I_1) \sin[2\pi(f_c + k f_1)t + k I_2 \sin(2\pi f_2 t)] \\ &= \sum_k \sum_n J_k(I_1) \cdot J_n(k I_2) \sin[2\pi(f_c + k f_1 + n f_2)t] \end{aligned}$$

Il risultato puo' venire interpretato come se ciascuna parziale prodotta dal modulatore f_1 sia a sua volta modulata da f_2 con indice di modulazione $k I_2$. Pertanto risulteranno le componenti di frequenza $|f_c \pm k f_1 \pm n f_2|$ con approssimativamente $0 \leq k \leq I_1$ e $0 \leq n \leq I_1 \cdot I_2$. La frequenza massima e'

$f_c + I_1(f_1 + I_2 f_2)$. La struttura dello spettro e' simile a quella prodotta da due modulanti sinusoidali, ma con banda maggiore. Anche qui se i rapporti sono semplici lo spettro sara' del tipo $|f_c \pm k f_m|$ dove f_m e' il massimo comun divisore tra f_1 e f_2 .

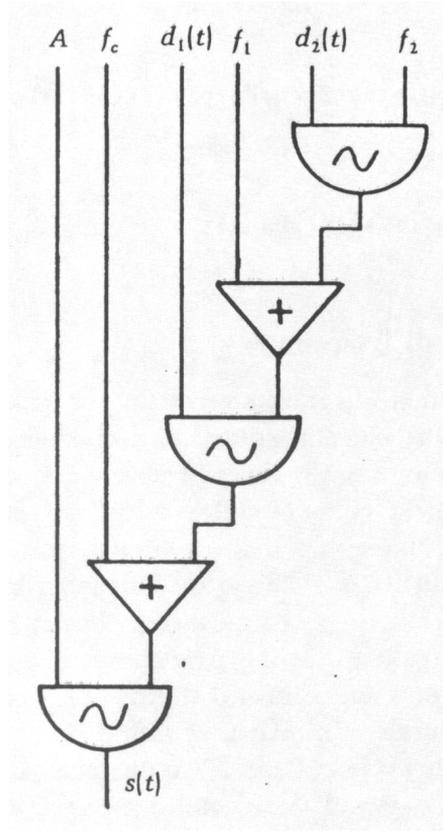


Figura 5.23: Modulazione di frequenza con due modulanti in cascata

Feedback FM

Consideriamo infine il caso in cui si usi come modulante il valore precedente del segnale generato. Si ha così la cosiddetta feedback FM. Essa è descritta in termini digitali da queste relazioni:

$$\begin{aligned}\phi(n) &= \beta s(n-1) \\ s(n) &= \sin\left(2\pi \frac{f_c}{F_s} n + \phi(n)\right)\end{aligned}$$

dove β è il fattore di feedback e agisce come fattore di scala o indice di modulazione per il feedback. Al crescere di β il segnale passa da sinusoidale verso la forma d'onda a dente di sega in modo continuo. Lo spettro è armonico di frequenza f_c con aumento graduale del numero di armoniche. In termini di funzioni di Bessel risulta

$$s(t) = \sum_k \frac{2}{k\beta} J_k(k\beta) \sin(2\pi k f_c t)$$

Esempio di implementazione di modulazione di frequenza in Matlab

Per concludere, si vuole ora tornare sull'osservazione fatta all'inizio di questa sezione riguardo l'implementazione dell'oscillatore modulato in frequenza. Si è detto che risulta immediato realizzare l'oscillatore trasformando la modulazione di frequenza in modulazione di fase. Tuttavia, in qualche occasione, si può volere l'imposizione della frequenza istantanea all'audio rate. È utile a questo punto chiarire come calcolare la fase della sinusoide da generare se, di volta in volta, la frequenza di controllo è costante, imposta al frame rate o imposta all'audio rate.

- f costante. In questo caso, ad ogni campione successivo, la fase cresce in modo lineare con il tempo ed è possibile calcolare il valore di fase ad ogni istante di campionamento a partire dal valore di fase al campione precedente:

$$\phi(n+1) = \phi(n) + \frac{2\pi f}{F_s}$$

- f imposta a frame rate. Sia f_0 la frequenza imposta al frame K , f_1 la frequenza imposta al frame $K+1$ e sia $f(n)$ la frequenza istantanea. Si fa l'ipotesi che nel frame la frequenza cambi linearmente nel tempo da f_0 a f_1 . Se si scrive $f(n) = f_0 + (f(n) - f_0)$, si ha

$$\phi(n+1) = \phi(n) + \frac{2\pi f_0}{F_s} + \frac{2\pi(f_1 - f_0)}{2Spf} \left(\frac{1}{F_s}\right)^2$$

dove si è indicato con Spf il numero di campioni in un frame.

- f imposta ad audio rate. Sia in questo caso $f(n)$ la frequenza istantanea. L'aggiornamento della fase al campione successivo è allora

$$\phi(n+1) = \phi(n) + \frac{2\pi f(n)}{F_s} + \frac{2\pi(f(n+1) - f(n))}{2} \frac{1}{F_s}$$

Il corpo della funzione `FMosc(t0, a, f, phi0)`, con frequenza di controllo f ad audio rate, è illustrato di seguito:

```
function s=FMosc(t0,a,f,phi0)
(... )
lastfreq=f(1);
lastphase=0;
phase=zeros(1,SpF);
for (i=1:nframes)
    for(k=1:SpF)
        phase(k)=lastphase+2*pi*lastfreq/Fs+pi*(f((i-1)*SpF+k)-lastfreq)*(1/Fs);
        lastphase=phase(k);
        lastfreq=f((i-1)*SpF+k);
    end
    s(((i-1)*SpF+1):i*SpF)=a(i).*sin(phase);
end
```

Conclusioni sulla modulazione di frequenza

In conclusione la sintesi FM è un metodo molto versatile per produrre molti tipi di suono. D'altra parte non ci sono metodi chiusi che consentano di derivare in modo preciso i parametri di un modello

FM analizzando un suono dato. Per riprodurre dei suoni di strumenti musicali e' quindi preferibile usare altre tecniche come la sintesi additiva o la sintesi per modelli fisici.

Non evocando questa tecnica di sintesi nessuna esperienza musicale nell'esecutore, il controllo parametrico risulta poco intuitivo e richiede quindi una rilevante dose di esperienza specifica, caratteristica questa degli strumenti innovativi. Il suo principale punto di forza, ossia l'elevata dinamica timbrica legata a pochi parametri e a basso costo computazionale, sta perdendo progressivamente terreno nei confronti di altre tecniche di sintesi, piu' costose, ma controllabili in maniera piu' naturale e intuitiva. Il metodo conserva comunque la particolarita' di definire un suo spazio timbrico peculiare. Esso pur non prestandosi particolarmente alla simulazione di qualita' di suoni naturali, offre comunque un grande ventaglio di sonorita' originali di notevole interesse per la computer music.

5.3.2 Spettri di tipo $f_1 \pm kf_2$

Le seguenti considerazioni valgono per i suoni prodotti mediante traslazione di spettri armonici, cioe' per suoni composti da parziali di frequenza $f_1 \pm kf_2$ con $k = 0, 1, \dots$. Questi spettri sono caratterizzati dal rapporto f_1/f_2 . Nel caso delle modulazioni esso equivale al rapporto f_c/f_m tra frequenza della portante e della modulante. Quando il rapporto puo' essere rappresentato con una frazione irriducibile $f_1/f_2 = N_1/N_2$ con N_1 e N_2 interi primi tra loro, il suono risultante e' armonico, nel senso che tutte le componenti sono multiple intere di una fondamentale. La frequenza fondamentale risulta

$$f_0 = \frac{f_1}{N_1} = \frac{f_2}{N_2}$$

Si vede che in questo caso f_1 coincide con la N_1 -esima armonica, infatti

$$f_1 = N_1 f_0$$

$$f_2 = N_2 f_0$$

Se $N_2 = 1$, tutte le armoniche sono presenti e le componenti laterali con k negativo si sovrappongono a quello con k positivo. Se $N_2 = 2$, sono presenti solo le armoniche dispari e le componenti si sovrappongono ancora. Se $N_2 = 3$, mancano le armoniche multiple di 3. In generale il rapporto N_1/N_2 e' anche un indice dell'armonicita' dello spettro. Intuitivamente il suono e' piu' armonioso, quando il rapporto N_1/N_2 e' semplice ossia quanto piu' il prodotto $N_1 \cdot N_2$ e' piccolo.

I rapporti possono essere raggruppati in famiglie. Tutti i rapporti del tipo $|f_1 \pm kf_2|/f_2$ possono produrre le stesse componenti del rapporto f_1/f_2 . Cambia solo quale parziale coincide con f_1 . Ad esempio i rapporti $2/3, 5/3, 1/3, 4/3, 7/3$ e cosi' via appartengono alla stessa famiglia. Sono presenti tutte le armoniche ad esclusione di quelle multiple di 3 (essendo $N_2 = 3$) e f_1 coincidera' rispettivamente con la seconda, quinta, prima, quarta e settima armonica. Il rapporto che distingue la famiglia si dice in forma normalizzata se e' minore o uguale a $1/2$. Nell'esempio precedente esso e' uguale a $1/3$. Ciascuna famiglia e' quindi caratterizzata da un rapporto in forma normalizzata. Spettri simili possono essere ottenuti da suoni della stessa famiglia. Si vede quindi che il denominatore N_2 e' caratterizzante lo spettro. In particolare per $N_2 < 5$ ogni denominatore definisce una sola famiglia.

Se il rapporto e' irrazionale, il suono risultante non e' piu' periodico. Questa possibilita' viene usata per creare facilmente suoni inarmonici. Ad esempio se $f_1/f_2 = 1/\sqrt{2}$ la spettro consiste in componenti a frequenza $f_1 \pm k\sqrt{2}$. Non c'e' quindi nessuna fondamentale implicita. Un comportamento simile si ottiene per rapporti non semplici come $f_1/f_2 = 5/7$.

Di particolare interesse e' il caso in cui il rapporto f_1/f_2 approssimi un semplice valore, cioe'

$$\frac{f_1}{f_2} = \frac{N_1}{N_2} + \epsilon.$$

In questo caso il suono non e' piu' rigorosamente periodico. La fondamentale e' ancora $f_0 = f_2/N_2$ e le parziali sono spostate dal loro preciso valore di $\pm \epsilon f_2$. Pertanto un piccolo spostamento della portante non cambia l'altezza del suono e lo rende molto piu' vivo grazie ai battimenti tra le componenti vicine. Si noti invece che lo stesso spostamento della modulante f_2 cambia la fondamentale.

5.3.3 Sintesi moltiplicativa

La trasformazione non lineare piu' semplice consiste nella moltiplicazione di due segnali. Nel campo analogico e' chiamata modulazione ad anello (*ring modulation*) o RM ed e' piuttosto difficile da produrre in modo preciso. Nel campo numerico invece consiste in una semplice operazione di moltiplicazione. Se $x_1(t)$ e $x_2(t)$ sono due segnali il segnale di uscita e' dato da

$$s(t) = x_1(t) \cdot x_2(t) \quad (5.14)$$

Lo spettro risultante e' dato dalla convoluzione tra gli spettri dei due segnali.

Normalmente uno dei due segnali e' sinusoidale di frequenza f_c ed e' chiamato portante $c(t)$ (*carrier*) e l'altro e' un segnale in ingresso alla trasformazione ed e' chiamato modulante $m(t)$. Si ha pertanto

$$s(t) = m(t) \cdot c(t) = m(t) \cos(2\pi f_c t + \phi_c)$$

e lo spettro risultante e'

$$S(f) = \frac{1}{2} [M(f - f_c)e^{j\phi_c} + M(f + f_c)e^{-j\phi_c}]$$

Lo spettro di $s(t)$ e' composto da due copie dello spettro di $m(t)$: una banda laterale inferiore (LSB) e la banda laterale superiore (USB). La LSB e' rovesciata in frequenza e entrambe le bande sono centrate attorno a f_c . A seconda della larghezza di banda di $m(t)$ e della frequenza della portante f_c , le bande laterali possono essere parzialmente riflesse attorno all'origine dell'asse di frequenza. Se la portante ha diverse componenti spettrali, lo stesso effetto si ripete per ogni componente. L'effetto acustico della modulazione ad anello e' relativamente facile da capire per segnali semplici. Diventa pero' piuttosto complicato da immaginare per segnali con numerose parziali. Se sia la portante che la modulante sono sinusoidali di frequenza rispettivamente f_c e f_m , si sente la somma di due differenti parziali a frequenza $f_c + f_m$ e $f_c - f_m$. Ad esempio se $f_c = 500$ Hz e $f_m = 400$ Hz, la modulazione ad anello produce due parziali a frequenza 900 Hz e 100 Hz. Se invece se $f_c = 100$ Hz e di nuovo $f_m = 400$ Hz, si producono due parziali a frequenza 500 Hz e -300 Hz. Quest'ultima ha frequenza negativa; si ha quindi una riflessione (*foldunder*) attorno allo 0 con cambio di segno della fase. Infatti $\cos(-2\pi 100t + \phi) = \cos(2\pi 100t - \phi)$. In definitiva si sentiranno due componenti a frequenza 500 Hz e 300 Hz.

Se la portante e' sinusoidale e la modulante e' periodica di frequenza f_m con $m(t) = \sum_{k=1}^N b_k \cos(2\pi k f_m t + \phi)$ risulta

$$s(t) = \sum_{k=1}^N \frac{b_k}{2} [\cos[2\pi(f_c + k f_m)t + \phi_k] + \cos[2\pi(f_c - k f_m)t - \phi_k]] \quad (5.15)$$

L'armonica k -esima dara' luogo a due righe, una nella LSB e l'altra nella USB, a frequenza $f_c - k f_m$ e $f_c + k f_m$. Lo spettro risultante ha quindi righe a frequenza $|f_c \pm k f_m|$ con $k = 1, 2, \dots$, dove si e' usato il valore assoluto per tenere conto delle possibili riflessioni attorno allo 0. Valgono per questi spettri le considerazioni fatte sopra sulle famiglie di spettri $|f_c \pm k f_m|$.

Modulazione di ampiezza

La modulazione di ampiezza era piu' facile da realizzare nel campo analogico e pertanto e' stata usata per molto tempo. Essa puo' essere implementata come

$$s(t) = [1 + \delta m(t)]c(t) \quad (5.16)$$

dove si e' assunto che l'ampiezza di picco di $m(t)$ sia 1. Il coefficiente δ determina la profondita' di modulazione. L'effetto e' massimo quando $\delta = 1$ e viene disattivato quando $\delta = 0$.

Tipiche applicazioni sono l'uso di un segnale audio come portante $c(t)$ e un oscillatore a bassa frequenza (LFO) come modulatore $m(t)$. L'ampiezza del segnale audio varia seguendo l'ampiezza di $m(t)$ e cosi' verra' sentita. Se il modulatore e' un segnale udibile e la portante una senoide di frequenza f_c , l'effetto e' simile a quello visto per il modulatore ad anello, solo che in uscita si sentira' anche la frequenza della portante f_c .

Si noti che a causa del tempo di integrazione del nostro sistema uditivo, l'effetto e' percepito diversamente in dipendenza del campo di frequenza dei segnali considerati. Una modulazione con frequenza sotto 20 Hz sara' sentita nel dominio del tempo (variazione di ampiezza), mentre modulazioni con frequenza superiori verranno sentite come componenti spettrali distinte (banda laterale inferiore, portante, banda laterale superiore).

5.3.4 Sintesi per distorsione non lineare

L'idea fondamentale della sintesi per distorsione non lineare, conosciuta anche sotto il nome di *waveshaping* e' quella di passare una senoide per un blocco distorcente. E' noto infatti che se una senoide passa per filtro lineare viene modificata la sua ampiezza e fase, ma non la forma d'onda. Se invece l'amplificatore e' non lineare la forma d'onda del segnale viene modificata e vengono create altre componenti spettrali. Questo fatto e' ben noto nei segnali analogici, dove si cerca di evitarlo o usarlo per creare effetti tipo amplificazione con tubi elettronici. Nel campo digitale si e' pensato di sfruttarlo per produrre suoni periodici di spettro variabile. Il blocco distorcente e' realizzato mediante una funzione non lineare $F(x)$ chiamata funzione distorcente o *shaping function* memorizzata su tabella. Piu' raramente la funzione viene calcolata direttamente. La funzione distorcente dipende solo dal valore istantaneo dell'ingresso. Pertanto in corrispondenza ad un segnale di ingresso $x(t)$ il metodo calcola

$$s(t) = F[x(t)] \quad (5.17)$$

cercando in tabella ad ogni campione il valore all'ascissa $x(t)$.

Questa tecnica puo' essere usata come effetto audio che oer la sintesi. Nel primo caso si usa una leggera distorsione, spesso sotto forma di saturazione, su un segnale qualsiasi per arricchire un po' lo spettro e simulare l'effetto che si verifica sovente in strumenti meccanici o elettronici analogici.

Per la sintesi dei suoni normalmente si usa un ingresso sinusoidale di ampiezza I (che puo' essere variata)

$$x(t) = I \cdot \cos(2\pi ft)$$

per cui la formula di sintesi diviene:

$$s(t) = F[x(t)] = F[I \cdot \cos(2\pi ft)]$$

In figura 5.24 e' riportato lo schema a blocchi della sintesi per distorsione non lineare. Con il parallelogramma viene indicato il modulo che effettua la distorsione mediante lettura da tabella del valore di $F(x)$.

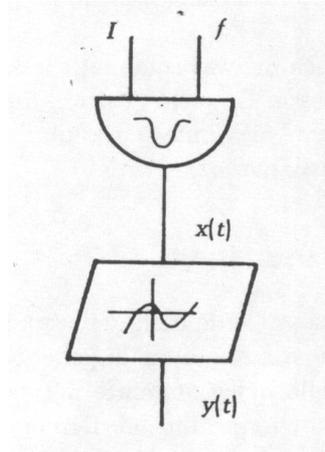


Figura 5.24: Sintesi per distorsione non lineare

In generale nella sintesi, se $F(x) = F_1(x) + F_2(x)$, la distorsione prodotta da $F(\cdot)$ è uguale alla somma di quelle prodotte da $F_1(\cdot)$ e $F_2(\cdot)$ separatamente. In particolare una funzione pari, cioè simmetrica rispetto all'asse y genera solo armoniche pari e una funzione dispari (antisimmetrica) genera solo armoniche dispari. Normalmente una funzione distorcente produce infinite armoniche. Se però la funzione è un polinomio $p(x)$ di grado N , vengono prodotte solo le prime N armoniche. In questo modo si può controllare il foldover. Se la funzione è polinomiale è anche facile calcolare le ampiezze delle armoniche generate dalla distorsione di una sinusoidale. Infatti, tenendo conto che il polinomio di Chebyshev di ordine k è definito come $T_k(\cos \theta) = \cos(k \cdot \theta)$, ne deriva che usandolo come funzione distorcente di una sinusoidale di ampiezza unitaria si ha

$$s(t) = F[I \cdot \cos(2\pi ft)] = T_k[\cos(2\pi ft)] = \cos(2\pi kft)$$

Viene cioè generata l'armonica k -esima. Pertanto, sviluppando il polinomio distorcente $p(x)$ in serie di polinomi di Chebyshev

$$F(x) = p(x) = \sum_{i=0}^N d_i x^i = \sum_{k=0}^N h_k T_k(x)$$

si ottengono le ampiezze delle armoniche, dove h_k sarà l'ampiezza dell'armonica k -esima. Usando quindi il polinomio $p(x)$ così determinato come funzione distorcente di una sinusoidale ad ampiezza $I = 1$, si ottiene

$$s(t) = F[\cos(2\pi ft)] = \sum_{k=0}^N h_k \cos(2\pi kft)$$

Se varia l'ampiezza dell'ingresso I , varia anche la distorsione e lo spettro prodotto. Questo è simile ad un'espansione o contrazione della funzione, in quanto viene usata una parte maggiore o minore della funzione. L'ampiezza e l'intensità del suono prodotto varia quindi con l'ampiezza dell'ingresso e deve essere pertanto compensata con un'opportuna funzione di normalizzazione. Nel caso invece che la distorsione sia usata come effetto, questi cambiamenti spesso sono in accordo con il fenomeno acustico che si vuole imitare e quindi non devono essere compensati.

Un'altra variazione dinamica del waveshaping che è facile implementare consiste nell'aggiungere una costante all'ingresso. In questo caso la funzione viene traslata orizzontalmente. Anche in questo caso lo spettro varia, ma non è più separabile l'effetto della parte pari da quella dispari della funzione originaria.

5.4 Sintesi per modelli fisici

Gli algoritmi visti sopra si basano su modelli del segnale che arriva alle nostre orecchie. Essi appartengono quindi alla categoria dei modelli generativi. La sintesi per modelli fisici segue invece un approccio alternativo, in cui si cerca di rappresentare la dinamica degli oggetti (reali o virtuali) responsabili della produzione del suono. La sintesi è quindi basata sull'uso di modelli formali di strumenti musicali tradizionali; il suono viene generato simulando numericamente la dinamica dello strumento che lo produce.

I *modelli del segnale* hanno fino ad oggi dominato la scena, in virtù di algoritmi efficienti e flessibili. I *modelli fisici* di strumenti musicali, sebbene costituiscano da sempre un campo di indagine vivo e in evoluzione, hanno invece cominciato solo di recente ad essere usati per la sintesi del suono. I motivi sono da ricercarsi principalmente in una maggiore complessità degli algoritmi e nello stesso tempo in una più marcata rigidità delle strutture di sintesi; la maggior parte dei modelli rappresenta infatti strumenti specifici, e raramente è possibile trovare delle strutture che descrivano ampie classi di strumenti musicali in maniera efficiente ed efficace.

D'altro canto, la sintesi per modelli fisici presenta alcune notevoli potenzialità. Innanzitutto, si può osservare che essa impiega algoritmi caratterizzati da pochi parametri di controllo, i quali hanno quasi sempre un significato fisico. Da ciò segue che poche ed intuitive variazioni sui parametri stessi producono dei risultati sonori coerenti con l'esperienza, anche gestuale, del musicista. In altre tecniche basate su modelli del segnale questo non accade, proprio perché tali modelli si disinteressano dei meccanismi di generazione del suono. Questa proprietà è particolarmente interessante dal punto di vista musicale; si può infatti pensare al controllo dello strumento simulato in termini degli stessi parametri usati per il controllo del corrispondente strumento reale; tali parametri possono essere forniti dall'esecutore attraverso opportuni trasduttori ed attuatori, che permettano di riprodurre interventi gestuali che fanno parte dell'esperienza del musicista.

Per quanto riguarda le sonorità prodotte, queste sono caratterizzate da un'evoluzione timbrica naturale, analoga a quella degli strumenti acustici. In particolare, vengono conservate quelle caratteristiche di *varietà* e di *coerenza* timbrica tipiche degli strumenti acustici e gradite all'orecchio umano.

5.4.1 Blocchi funzionali

Nello sviluppo dei modelli è utile individuare blocchi funzionalmente distinti; ciascuno di essi può essere studiato in maniera indipendente, e la dinamica complessiva del sistema viene data dall'interazione delle parti.

Per gli strumenti musicali, un primo livello di scomposizione è dato dall'identificazione di due blocchi funzionali distinti: l'*eccitatore* e il *risonatore*. Il risonatore è la parte dello strumento in cui la vibrazione ha effettivamente luogo, ed è correlato a caratteristiche sonore quali altezza ed inviluppo spettrale. L'eccitatore provoca ed eventualmente sostiene la vibrazione nel risonatore, immettendo energia nello strumento; da esso dipendono le proprietà di attacco del suono, fondamentali nell'identificazione del timbro. Per fare qualche esempio, sono blocchi risonatori la corda nella chitarra, nel pianoforte, nel violino, o il tubo acustico nei legni e negli ottoni. Sono invece eccitatori il plettro nella chitarra, l'archetto nel violino, il martelletto nel pianoforte, l'ancia nel clarinetto. L'interazione può essere semplicemente *feedforward*, se l'eccitatore non riceve nessuna informazione di ritorno dal risonatore, oppure *feedback*, se i due blocchi si scambiano informazione in entrambe le direzioni. Nella chitarra il meccanismo di eccitazione è con buona approssimazione *feedforward*: il plettro pizzica

la corda fornendo le condizioni iniziali, poi il risonatore evolve in maniera libera. Strutture di tipo feedback sono invece adatte per modellizzare strumenti persistentemente eccitati, come i fiati.

La suddivisione in blocchi può essere estesa a livelli più fini, soprattutto per i risonatori. In una chitarra, il blocco del risonatore comprenderà la corda vibrante e la tavola armonica; in uno strumento a fiato, invece, esso sarà composto da dal tubo acustico, dalla campana (la parte terminale) e dai fori. Ad ogni stadio ci si trova davanti ad una scelta: continuare nel raffinamento, modellizzando la fisica del sistema fino ai livelli più bassi (modelli *white box*), oppure considerare solo le proprietà ingresso-uscita del blocco in esame (modelli *black box*). È chiaro che un modello *black box* risulta più semplice nella progettazione e più efficiente nell'implementazione, riducendo il blocco funzionale ad un semplice filtro. Si rischia tuttavia di introdurre semplificazioni eccessive nel modello, e comunque di produrre strutture i cui parametri non hanno significato fisico e sono quindi difficilmente controllabili. Dalla parte opposta, un approccio completamente *white box* genera modelli che possono rivelarsi eccessivamente pesanti dal punto di vista computazionale.

Passando dalla scomposizione concettuale allo sviluppo dei modelli, i due principali blocchi dello strumento sono rappresentati da altrettanti sistemi dinamici. L'eccitatore contiene nella maggior parte dei casi delle forti non linearità; i risonatori, descritti tipicamente dall'equazione delle onde, vengono invece modellizzati da sistemi lineari. Nelle implementazioni numeriche si deve prendere in considerazione un terzo elemento, un modulo di interazione tra i due blocchi. Lo schema risultante è allora quello di figura 5.25.



Figura 5.25: Schema di interazione tra eccitatore e risonatore.

Il modulo di interazione ha principalmente la funzione di adattatore, converte le variabili in arrivo da eccitatore e risonatore e permette di ovviare ad anomalie sorte durante il processo di discretizzazione (ad esempio la nascita di anelli senza ritardi e quindi non computabili). Questo blocco può anche essere usato per compensare alcune semplificazioni del modello; ad esempio, nel modello di un clarinetto si può introdurre nel blocco di interazione un rumore fluidodinamico, la cui ampiezza è modulata dal flusso d'aria.

5.4.2 Elementi concentrati e distribuiti

Nel costruire un modello a tempo continuo per un sistema acustico, due sono le strade che si possono seguire. Spesso è possibile utilizzare una schematizzazione che rappresenta l'oggetto in esame come costituito da corpi rigidi connessi tra loro attraverso degli ideali *elementi concentrati*: molle, elementi di attrito, e così via. Il modello così ottenuto è allora descritto da sistemi di equazioni differenziali ordinarie. Alternativamente, l'oggetto in esame può venire trattato come un corpo flessibile, in cui le forze e la materia sono distribuite in uno spazio continuo; si parla allora di *elementi distribuiti*, come corde, o membrane, o ancora l'aria all'interno di un tubo acustico. In questo caso il modello viene descritto da equazioni alle derivate parziali, nelle quali le quantità fisiche sono funzioni anche della posizione, oltre che del tempo.

In entrambi i casi, le equazioni usate introducono approssimazioni e semplificazioni della realtà fisica. Un secondo livello di approssimazione viene poi dalla simulazione, dove le equazioni di partenza devono essere risolte numericamente.

Per quanto riguarda la simulazione di equazioni differenziali ordinarie, e quindi di modelli ad elementi concentrati, il calcolo numerico mette a disposizione una varietà di metodi: schemi di Eulero in avanti o all'indietro, trasformazione bilineare, metodi lineari multistep, metodi di Runge-Kutta, solo per nominare i principali.

Uno dei metodi più popolari per la risoluzione numerica di equazioni alle derivate parziali è quello che va sotto il nome di *differenze finite*; in esso i domini spaziale e temporale sono ricoperti da griglie, e le derivate vengono sostituite da combinazioni lineari dei valori della variabile sui punti della griglia. I principali problemi introdotti dagli schemi alle differenze finite riguardano innanzitutto la stabilità delle soluzioni discrete, ed in secondo luogo l'introduzione di dissipazione e dispersione numeriche. Si pensi ad esempio all'equazione delle onde, valida in una corda o un tubo cilindrico ideali:

$$\frac{\partial^2 y}{\partial t^2}(x, t) = c^2 \frac{\partial^2 y}{\partial x^2}(x, t), \quad (5.18)$$

in cui c è la velocità di propagazione, t ed x sono rispettivamente le variabili temporale e spaziale, e y è ad esempio la velocità trasversale della corda o la pressione nel tubo.

È ben noto che le soluzioni della (5.18) sono forme d'onda che traslano rigidamente con velocità c . La dissipazione numerica introduce allora delle attenuazioni (o delle amplificazioni) artificiali nelle soluzioni approssimate, mentre la dispersione cambia la velocità di traslazione.

Nell'ambito delle equazioni *iperboliche*, di cui la (5.18) è un esempio, esistono metodi standard che permettono di valutare l'efficacia di un metodo alle differenze finite in rapporto a questi problemi. Un primo risultato è fornito dalla *condizione di Courant-Friedrichs-Lewy (CFL)*, che è una condizione necessaria (e non sufficiente) per la stabilità di uno schema numerico alle differenze finite; nel caso dell'equazione delle onde la condizione assume il seguente aspetto:

$$c \frac{\Delta t}{\Delta x} \leq 1 \quad (5.19)$$

La condizione *CFL* stabilisce dunque una relazione tra i passi di integrazione in tempo e spazio, ed ha come interessante (e non intuitiva) conseguenza che Δx troppo piccoli non aumentano l'accuratezza del metodo: al contrario, possono portare ad instabilità. Nel caso in cui c non sia una costante ma una funzione di spazio e tempo la condizione rimane valida, a patto di sostituire la costante c con $\sup_{x,t} c(x, t)$.

Un potente strumento per lo studio delle proprietà di un metodo alle differenze finite è fornito dall'*analisi di Von Neumann*. Questa utilizza gli sviluppi in serie di Fourier per trovare dei coefficienti di amplificazione su ciascuna armonica; lo studio di tali coefficienti permette di imporre condizioni su Δx e Δt sufficienti per la stabilità del metodo, e mette allo stesso tempo in luce le proprietà di dispersione e dissipazione del metodo stesso.

Oltre ai modelli alle differenze finite esaminati fino ad ora, esistono altre tecniche per modellizzare sistemi distribuiti. Un possibile approccio alternativo è quello di scomporre il sistema in un insieme di particelle interagenti, modellizzando la dinamica delle particelle e definendo le leggi di interazione tra esse. Si parla allora di *modelli cellulari*. L'esempio più significativo fornito in questa direzione è il sistema *CORDIS-ANIMA*, introdotto da Claude Cadoz ed altri all'inizio degli anni '90. Questo modello fa riferimento ad una sequenza di masse elementari, collegate tra loro da masse ed attriti, e si presta a simulare efficacemente corpi vibranti quali lastre, sbarre, corde o membrane. Per contro,

presenta almeno due svantaggi: innanzitutto non esistono strumenti analitici (come l'analisi di Von Neumann nei modelli alle differenze finite) per valutare la correttezza della discretizzazione, ed in particolare per individuare dei vincoli di stabilità numerica. In secondo luogo, i costi computazionali sono molto elevati; questo perché nel modello si prende in considerazione la dinamica di tutti i punti dell'oggetto, fornendo informazione ridondante rispetto a quella musicalmente importante.

Un altro importante approccio nella modellizzazione di sistemi vibranti è quello fornito dai *modelli waveguide*, sviluppati principalmente da J.O. Smith. Il punto di partenza di tali modelli consiste nel discretizzare le soluzioni dell'equazione delle onde piuttosto che l'equazione stessa. Questo porta a delle strutture estremamente semplici ed efficienti, in grado di simulare efficacemente corde e tubi acustici. Entro certi limiti è possibile riprodurre anche fenomeni di dissipazione e di dispersione tipici dei sistemi reali.

5.4.3 Il ruolo delle non linearità

La presenza di relazioni non lineari gioca un ruolo essenziale nello sviluppo di modelli di sistemi acustici; in assenza di esse, questi si ridurrebbero a schemi di sintesi sottrattiva, in cui un segnale di eccitazione viene sottoposto a filtraggio. Come già detto le non linearità sono presenti in particolare negli eccitatori e nei meccanismi di interazione di questi con i risonatori; poiché l'area in cui l'eccitazione ha luogo è solitamente piccola, è naturale utilizzare modelli concentrati per rappresentarle. Ad un primo livello di approssimazione si può supporre che il meccanismo di eccitazione sia rappresentato da una funzione non lineare istantanea, la cui forma viene a volte dedotta da osservazioni sperimentali e direttamente implementata nella simulazione. Un esempio è dato da un modello di corda strofinata in cui, per diversi valori della pressione e della velocità dell'archetto (i parametri di controllo), la velocità trasversale della corda e la forza di frizione tra corda ed archetto sono legate da una relazione altamente non lineare. Un'analoga relazione si trova tra pressione e flusso in un modello semplificato di ancia di clarinetto.

Per realizzare simulazioni più accurate, è necessario passare da una rappresentazione mediante non linearità istantanee ad una che tenga in considerazione la dinamica dell'eccitatore. Modelli dinamici di questo tipo esistono ad esempio per l'ancia di clarinetto, o per l'interazione martelletto-corda nel pianoforte. Un problema che allora sorge è come trasportare tali modelli dinamici e non lineari in schemi computazionali efficienti. Una struttura utilizzata con successo in molti casi è quella illustrata in figura 5.26.

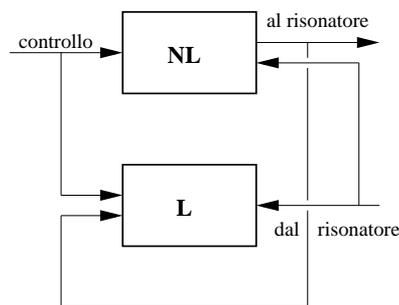


Figura 5.26: Schema generale per un eccitatore dinamico.

Il modello ad elementi concentrati dell'eccitatore viene suddiviso in due blocchi: il primo, contrassegnato con la lettera L, contiene la dinamica del sistema ed è lineare; il secondo, NL, è una funzione

non lineare istantanea delle variabili provenienti dall blocco L, e dipende anch'esso dai parametri di controllo.

Quando si discretizza un tale modello, ad esempio con una tecnica alle differenze finite, l'anello di feedback tra i due blocchi produce tipicamente loop senza ritardi e quindi non computabili. Questo inconveniente può essere superato introducendo nell'anello dei ritardi fittizi, ma in tal modo si introducono errori in molti casi non accettabili. Metodi più raffinati per risolvere le non computabilità sono ad esempio il *metodo W* e il *metodo K*. Il primo è un metodo locale, nel senso che mediante l'uso di variabili d'onda permette di "adattare" ciascun elemento, in maniera tale da rendere computabili gli schemi. Il secondo è invece globale: utilizzando il teorema di funzione implicita, permette di passare a delle nuove variabili che risultano computabili ad ogni passo.

5.5 Modelli per la sintesi dei segnali di controllo

Nei paragrafi precedenti si sono visti alcuni dei più importanti modelli per la sintesi del suono. Essi costituiscono il meccanismo sottostante alla produzione del suono sintetico.

Il problema del controllo nella sintesi fa riferimento a tutto quello che è necessario per passare dalla descrizione simbolica dei suoni, come espressa nella partitura, al suono, utilizzando i modelli di sintesi. Tradizionalmente la partitura prevede un insieme di note (simboli che descrivono un suono e le sue proprietà a livello astratto) ed è compito dell'esecutore, con l'aiuto di uno strumento, di tradurle in suono. In generale possono essere distinti due livelli di astrazione nel controllo, cui corrispondono scale dei tempi diverse:

- controllo dell'espressività dello strumento
- controllo della dinamica spettrale

Il primo, che riguarda l'esecutore inteso come interprete, fa riferimento al passaggio da simboli ad azioni nel tempo per scegliere e rendere gli effetti espressivi voluti. In genere esso non rappresenta una semplice trasformazione di simbolo in simbolo, ma piuttosto determina la variazione continua di un insieme di parametri. Esso consiste quindi nella generazione di segnali che variano sulla scala dei tempi delle frasi. Con queste azioni il musicista dirige e dà forma al fluire del suono musicale che costituisce l'opera.

Il secondo livello invece controlla la dinamica spettrale di una nota e determina il passaggio dai parametri espressivi all'algoritmo sottostante. In questo caso i segnali variano durante l'evoluzione della nota e operano sulla scala di tempo della sua durata. Al primo livello è associata la proprietà di "suonabilità", ossia la possibilità data all'esecutore di interagire in maniera soddisfacente con lo strumento; al secondo livello è associata l'idea di qualità del timbro, ossia la capacità dello strumento di produrre "bei suoni". In entrambi i casi l'obiettivo è comunque quello di estrarre o sintetizzare un insieme di segnali di controllo più semplici da gestire.

Tra esecutore e strumento tradizionale esiste un'interfaccia, ad esempio tastiera o archetto, che determina e vincola le possibili azioni. Tra l'algoritmo di sintesi e il musicista è pure presente un'interfaccia di controllo: essa costituisce ciò che il musicista conosce dello strumento e l'oggetto con cui interagisce. L'interfaccia di controllo mappa l'insieme delle possibili azioni dell'esecutore nei parametri del controllo degli algoritmi sottostanti in maniera tale da presentare una coerenza e una consequenzialità che soddisfi le aspettative del musicista. Negli strumenti commerciali, tale interfaccia è determinata dal costruttore. Invece l'uso del computer ha reso le interfacce programmabili a seconda delle esigenze dell'utente, per cui esse possono essere a diversi livelli di astrazione. Si

puo' andare da un controllo dettagliato dei parametri degli algoritmi di sintesi, come puo' essere per i musicisti utilizzando programmi tipo MusicV, all'esecuzione automatica di una partitura, in cui tutto il controllo e' affidato a processi automatici.

I segnali di controllo sono caratterizzati da alcune differenze fondamentali rispetto ai segnali acustici. In particolare, per la maggior parte di essi non appare significativa l'interpretazione in frequenza, mancando il segnale delle caratteristiche di periodicit  o formantizzazione tipiche del segnale acustico. Sembra piu' adeguata l'interpretazione temporale e quindi le tecniche di analisi, sintesi e manipolazione che fanno riferimento al tempo. Tuttavia, e' possibile utilizzare le tecniche viste sopra, opportunamente reinterpretate, per sintetizzare segnali generici, come quelli di controllo.

5.5.1 Riproduzione

Vi sono alcuni modelli di sintesi del suono che hanno degli algoritmi di analisi sufficientemente accurati. Ad esempio i metodi che adottano la rappresentazione tempo-frequenza, visti nella sintesi additiva, possiedono algoritmi come la STFT che consentono di stimare, a partire da un suono reale, i parametri del modello per una riproduzione accurata del suono originale. Questi parametri, come visto, sono i segnali di controllo della frequenza e dell'ampiezza di ogni parziale del suono esaminato. In questo modo si ottengono da un unico segnale a frequenza acustica, vari segnali di controllo che variano lentamente nel tempo. L'interpretazione del loro andamento ora viene fatta essenzialmente nel tempo, come variazione delle caratteristiche timbriche del suono.

Questi segnali di controllo vengono quindi spesso elaborati con le tipiche tecniche nel tempo, tipo cut and paste, riscaldamenti di ampiezza o cambio della scala temporale etc. In ogni caso il punto di riferimento e' che essi esprimono in dettaglio la variabilit  presente nei suoni naturali. In particolare, quando essi derivano dall'analisi del controllo espressivo mantengono caratteristiche della gestualit  del performer. Senza un modello piu' preciso di questa variabilit  si ricorre alla riproduzione. Questo metodo e' l'analogo della tecnica di campionamento vista per i segnali audio. Se si controlla lo stesso parametro, si ha la risintesi, eventualmente modificata del suono originale; altrimenti si possono controllare altri parametri del suono, ad esempio con l'involuppo d'ampiezza controllare la bandwidth. Un uso piu' creativo del sampling e' l'impiego di segnali rilevati da un timbro per il controllo di un timbro diverso. Ad esempio, le deviazioni di frequenza di una voce possono essere usate per controllare il pitch di un violino; ne risulta quindi un suono con caratteristiche timbriche ibride tra i due strumenti (si percepisce un violino che canta).

5.5.2 Controlli composti

Spesso e' conveniente ricorrere a sovrapposizione o concatenazione di elementi semplici di controllo. Tipico e' il caso in cui si genera l'andamento deterministico del segnale e si sovrappongono delle variazioni aleatorie. Ad esempio se pensiamo al controllo della frequenza di un suono, si puo' sommare un segnale che fornisce l'andamento generale della frequenza al suo stabilirsi, con una modulazione periodica di alcuni Hz piu' una variazione aleatoria o frattale.

In un controllo timbrico inoltre si puo' produrre il segnale di controllo come successione di forme d'onda diverse per le diverse parti del suono, e per ogni parte si puo' scegliere tra un repertorio di andamenti tipici. Ad esempio, l'ADSR e' un modello del segnale di controllo comunemente usato per l'involuppo d'ampiezza, il controllo spettrale ecc. In esso vengono giustapposte quattro "fasi temporali" di evoluzione del segnale di controllo (Attacco, Decadimento, Sostenuto, Rilascio).

Interpolazione

Una tecnica di sintesi molto utilizzata nella sintesi di segnali di controllo e' quella dell'interpolazione, sia lineare che non lineare. Nell'interpretazione che vede il processo di sintesi come un passaggio da "poche" informazioni localizzate ad una variazione continua di "molti" parametri, l'interpolazione offre un quadro concettuale di riferimento molto efficace. Infatti, in base alla specifica di pochi punti significativi nel segnale (ad esempio, il pitch delle note che compongono una frase) il modello consente di ricavare un segnale caratterizzato da zone di transizione "smooth". L'attenzione nell'impiego di questo modello va quindi posta nella specifica del modo in cui queste interpolazioni vanno fatte.

5.5.3 Modelli stocastici

La riproduzione dei segnali di controllo soffre degli stessi problemi visti con la riproduzione del suono; in particolare necessita di memorizzare completamente le funzioni e ha poca versatilita'. Solo recentemente si comincia a rivolgere l'attenzione a questo tipo di segnali per vedere di caratterizzarli e porli in relazione con le proprieta' acustiche dei suoni controllati. Quando queste dipendono da molte cause concomitanti, spesso e' utile usare un modello che distingue un andamento medio, dalle variazioni che sono modellabili da un processo aleatorio. In quest'ultimo e' piu' importante riprodurre alcune proprieta' statistiche piu' che l'esatto andamento. In questi casi si puo' procedere alla sintesi mediante un rumore bianco filtrato da un opportuno filtro lineare (ARMA); i parametri del filtro possono essere stimati dall'analisi del segnale. Questo e' applicabile in generale in tutti i casi nei quali non si e' in grado di fare specifiche ipotesi sulla struttura del controllo, ma e' possibile stimare i parametri statistici dello stesso.

Modelli frattali del segnale

In alcuni casi le variazioni sono dovute a caoticita' nel generatore, come ad esempio nelle canne d'organo eccitate dal getto d'aria. Questa caoticita' si riflette sulla natura frattale dei segnali di controllo. Si possono quindi usare modelli frattali, stimando la dimensione frattale del segnale, eventualmente tempo variante, e poi utilizzarla nella sintesi, ad esempio controllando un generatore di Fractional Brownian Motion noise. Esistono vari tipi di questi generatori. In genere essi fanno riferimento alle caratteristiche spettrali del segnale che sono del tipo $1/f^\gamma$ dove γ e' un opportuno esponente legato alla dimensione frattale e cioe' alla turbolenza del segnale. Essi sono realizzati o mediante filtraggio di rumore bianco, oppure come somma di rumori filtrati con bande diverse [Voss]. Va detto che un segnale frattale non manifesta le sue caratteristiche se ci limita a tratti di breve durata, in quanto si perdono le relazioni sui tempi lunghi. Questo porta talvolta ad un uso improprio dei generatori di segnali frattali.

Sono molto usate anche le mappe iterate per produrre frattali, sull'esempio di quanto si fa nella computer graphics con gli insiemi di Mandelbrot, Julia etc. . I segnali cosi' prodotti non sono pero' altrettanto efficaci per il controllo dei parametri musicali. Altri algoritmi, come *mid-point displacement*, consentono l'interpolazione frattale fra punti dati, con controllo locale della frattalita'.

L'interesse dato dal modello frattale e' che questo coglie un'importante aspetto temporale dei segnali naturali e cioe' della autoaffinita', che e' la somiglianza statistica dell'andamento temporale se visto su scale diverse del tempo. In particolare con un unico generatore si hanno segnali che evolvono su piu' scale di tempo simultaneamente.

Si puo' dire che i modelli a filtraggio ARMA risultano piu' indicati per modellare l'autocorrelazione sulle scale di tempo brevi, mentre i segnali frattali si possono sfruttare anche sulle scale di tempo lunghe.

5.5.4 Modelli fisici

Una possibilita' e' di usare modelli fisici anche per la sintesi dei segnali di controllo. In questo caso il sistema sara' lentamente variabile e provvedera' esso stesso a fornira una dinamica per l'evoluzione del segnale. Questa possibilita' e' pero' finora scarsamente usata, in vista invece di un controllo diretto.

In alcuni casi si usa un approccio di questo tipo, non per modellare qualche meccanismo esistente, ma si usa l'analogia fisica, solo come metafora descrittiva di procedimenti mentali. Ad esempio Todd propone l'analogia delle accelerazioni di una pallina che corre su una superficie con vari buchi, per descrivere le accelerazioni e rallentamenti espressivi del tempo musicale. Oppure Sundberg e Verrillo propongono l'analogia del rallentamento finale di un pezzo musicale con l'arrestarsi del passo di una persona. Questi modelli introdurrebbero qualcosa che viene cognitivamente percepito come plausibile e conosciuto.

5.5.5 Sintesi basata sull'apprendimento

Nel caso si desideri sintetizzare un segnale appartenente ad una classe di cui sono noti esempi rilevanti ma non si desidera ricorrere ai metodi sopra esposti e' possibile impiegare modelli generali e stimarne i parametri mediante "apprendimento" sui segnali campione. Tipico esempio e' l'impiego di generatori basati su reti neurali feed-forward che se opportunamente addestrate dimostrano buone proprieta' di generalizzazione producendo quindi risultati plausibili anche in casi diversi da quelli previsti nell'addestramento e, in ogni caso, percepiti come appartenenti alla classe modellata. Ad esempio si sono ottenute dinamiche spettrali per mezzo dell'interpolazione di spettri statici, mediante reti neurali opportunamente addestrate.

5.5.6 Sistemi di regole

Fino ad ora si sono presi in esame metodi di sintesi basati sul modello del segnale. E' tuttavia possibile ricorrere a modelli del generatore del segnale, ossia a modelli del controllore. Nella situazione specifica, si tenta quindi di modellare il comportamento dell'esecutore umano; una soluzione comunemente impiegata e' quella dei generatori di segnali basati su regole. Questi ultimi presuppongono che sia possibile estrarre in forma euristica un complesso di "regole di comportamento" del generatore in funzione dei simboli di controllo. Queste regole possono essere fornite ad un sistema esperto, in grado di simulare il comportamento desunto dall'analisi delle performances degli esecutori "campione".

In alcune situazioni l'insieme di regole risulta caratterizzato da un grado di incertezza tale da non consentire l'implementazione di regole binarie; in questi casi, si rivelano utili i controllori basati sulla "fuzzy logic". In essi vengono specificati insiemi di regole basati su variabili linguistiche (es. "se la nota e' lunga...") e vengono specificate le azioni da intraprendere se le condizioni di appartenenza risultano soddisfatte (es. "... allungala ancora un po'"). Mediante "defuzzificazione" e' poi possibile ottenere i valori numerici necessari al controllo.

5.5.7 Conclusioni sulla sintesi dei segnali di controllo

I metodi visti rappresentano casi limite dei metodi impiegati nella pratica per la sintesi dei segnali di controllo. E' infatti assai comune trovare metodi ibridi, ottenuti dalla combinazione dei metodi sopra esposti, cercando di sfruttare le qualita' peculiari di ciascuna tecnica.

Si puo' osservare che la sintesi dei segnali di controllo usa modelli piuttosto semplificati. Per quanto riguarda il controllo espressivo, cio' e' dovuto al fatto che la problematica del controllo non

e' stata studiata in modo approfondito, anche per mancanza di adeguati strumenti di analisi. Inoltre, questo tipo di sintesi rappresenta il punto di incontro tra l'aspetto tecnico e l'aspetto artistico nell'informatica musicale, ed e' pertanto soggetto alle scelte e alle opinioni personali degli artisti. Per quanto riguarda il controllo della dinamica spettrale, ci sarebbero adeguati strumenti di analisi, ma e' mancato lo stimolo alla ricerca nel campo di nuovi modelli, principalmente perche' la qualita' dei suoni prodotti con i modelli semplificati e' spesso ritenuta piu' che soddisfacente. Cio' rappresenta un'indubbia conferma della validita' di tali metodi; tuttavia, si puo' osservare che lo sviluppo di modelli piu' pertinenti consentirebbe di lavorare ad un livello di astrazione superiore.

Indice

5 Sintesi dei segnali audio	5.1
5.1 Introduzione	5.1
5.1.1 Obiettivi della sintesi del suono	5.2
5.2 Metodi di generazione diretta	5.4
5.2.1 Generatori di forme d'onda	5.4
5.2.2 Campionamento	5.10
5.2.3 Sintesi additiva	5.11
5.2.4 Sintesi granulare	5.14
5.2.5 Sintesi sottrattiva	5.16
5.2.6 Sintesi della voce per predizione lineare	5.23
5.3 Sintesi non lineari	5.29
5.3.1 Sintesi per modulazione di frequenza	5.29
5.3.2 Spettri di tipo $f_1 \pm kf_2$	5.39
5.3.3 Sintesi moltiplicativa	5.40
5.3.4 Sintesi per distorsione non lineare	5.41
5.4 Sintesi per modelli fisici	5.43
5.4.1 Blocchi funzionali	5.43
5.4.2 Elementi concentrati e distribuiti	5.44
5.4.3 Il ruolo delle non linearità	5.46
5.5 Modelli per la sintesi dei segnali di controllo	5.47
5.5.1 Riproduzione	5.48
5.5.2 Controlli composti	5.48
5.5.3 Modelli stocastici	5.49
5.5.4 Modelli fisici	5.50
5.5.5 Sintesi basata sull'apprendimento	5.50
5.5.6 Sistemi di regole	5.50
5.5.7 Conclusioni sulla sintesi dei segnali di controllo	5.50