**Dispense di elaborazione analogica e numerica del segnale sonoro per la musica elettronica.**

# Analisi e sintesi

Molti modi di produrre, analizzare, trasformare i suoni guardandoli da diversi punti di vista.

# Indice generale

# 1 Note sul copyright



# 2 Introduzione

Questo testo si occupa di due opposte operazioni sui segnali: la sintesi e l'analisi.

Per *sintesi* si intende la creazione di segnali (più) complessi a partire da segnali costituenti (più) semplici prescelti come base, mediante un'operazione di "composizione" (o, brevemente, di *sintesi*). Per *analisi* si intende invece l'operazione di scomporre un segnale (più) complesso nei termini dei segnali (più) semplici prescelti, in modo che la corrispondente sintesi ricostruisca esattamente il segnale originario. Analisi e sintesi costituiscono quindi una *coppia di operazioni* la cui composizione (applicazione consecutiva) è l'identità, restituisce cioè i segnali di partenza.

Le definizioni qui adottate sono volutamente astratte, perché esiste un'infinità di modi di sintetizzare e di analizzare un segnale, specificamente sonoro. Ogni scelta di "segnali semplici" e di operazione di combinazione definisce una possibile coppia analisi-sintesi.

Tra tutti gli infiniti modi di concepire coppie analisi-sintesi, alcune rivestono un ruolo particolare, per ragioni vuoi storiche vuoi semantiche, intendendo con questo la circostanza che un determinato tipo di analisi-sintesi può riferirsi a particolari proprietà riscontrabili negli oggetti del mondo, o del nostro sistema uditivo.

Una particolare famiglia di analisi-sintesi è quella che adotta come segnali costituenti le sinusoidi (o cosinusoidi), e come operazione per la loro composizione la *combinazione lineare (miscelazione).* Con questo termine si indica in matematica la somma dei costituenti ciascuno moltiplicato per un suo coefficiente:

$$\mathrm{output}(t)=a_1\cdot \mathrm{s}_1(t)+a_2\cdot \mathrm{s}_2(t)+\ldots+a_n\cdot \mathrm{s}_n(t) \qquad 1$$

dove $\mathrm{s}_1(t)\ldots\mathrm{s}_n(t)$ sono i segnali costituenti, e $\mathrm{output}(t)$ è il segnale risultante dell'operazione di sintesi.

In questo contesto, dunque, *sintesi* indica l'operazione di produrre il segnale risultante effettuando una combinazione lineare dei segnali semplici, il che equivale a dire *scegliendo* i coefficienti $a_1\ldots a_n$ . *Analisi* indica invece l'operazione di individuare, a partire da un determinato segnale $\mathrm{output}(t)$ , i coefficienti $a_1\ldots a_n$ opportuni in grado di ricostruire esattamente il segnale di partenza mediante l'operazione di sintesi.

Se oltre ad adottare la combinazione lineare, scegliamo per $s_1(t) \dots s_n(t)$ delle sinusoidi, abbiamo la analisi-sintesi sinusoidale, che nel linguaggio della musica elettronica è correntemente chiamata *sintesi additiva* e *analisi additiva*.

Se i segnali sinusoidali di base vengono scelti con frequenze in rapporti armonici tra di loro, si parla di analisi e sintesi di Fourier.[1] Nella analisi-sintesi generica additiva dunque i segnali elementari presi in considerazione sono del tipo:

$$s_n(t) = \cos(\omega_n \cdot t + \varphi_n) \qquad 2$$

Nella analisi-sintesi di Fourier sono invece del tipo:

$$s_n(t) = \cos(n \cdot \omega_1 \cdot t + \varphi_n) \qquad 3$$

dove $\omega_1$ è la frequenza fondamentale di analisi-sintesi, e $n = 0, 1, 2, \dots$ .

Nel seguito, affronteremo prima la analisi-sintesi di Fourier, e poi quella generica, perché la seconda presenta maggiori difficoltà dal lato analisi, che verrà affrontata in una dispensa a parte, dedicata al *phase vocoder*. L'analisi di un segnale in termini sinusoidali di frequenza qualsivoglia prende il nome di analisi McAuley-Quatieri[2], la cui teoria è relativamente recente.

Altri metodi di analisi-sintesi sono quella per grani di Gabor (a cui idealmente si ispira la tecnica di composizione granulare), quella per ondine (*wavelet*), che è la base del metodo di compressione delle immagini "jpeg2000", molto più efficiente del più diffuso jpeg. Le ondine hanno avuto poca o nulla fortuna nelle applicazioni sonore o musicali.

La sintesi granulare, nel modo in cui è usata dai compositori contemporanei, è più un metodo compositivo che una tecnica di segnale, nel quale possono intervenire diverse e molteplici nozioni relative ai segnali, non specificamente legate a quella tecnica compositiva. La sintesi granulare non verrà pertanto trattata in questa dispensa.

## 3 Segnali periodici e quasi-periodici

Un segnale $s(t)$ si dice periodico, con periodo T, se, qualunque sia *t*, si ha

$$s(t+T) = s(t) \qquad 4$$

Questa definizione corrisponde al concetto intuitivo, ordinario, di periodicità: un segnale periodico si ripete identico a se stesso, dopo un periodo dato, per infinite volte. $T$ è, come detto, il *periodo* del segnale, e la sua *frequenza di ripetizione* è $\nu = 1/T$ .

E' facile convincersi tuttavia che se un segnale è periodico con periodo T, esso è periodico anche con periodo $2T, 3T, \dots nT \dots$ . Come periodo si intende quindi *il minore dei periodi ammissibili*.

1 Dal nome del matematico Joseph Fourier (1768-1830), che per primo adottò la corrispondente analisi per risolvere il problema matematico della propagazione del calore.

2 Questo tipo di analisi è detta "Analisi MQ" o "analisi sinusoidale", ed è stata definita, nell'ambito di ricerche sulla sintesi e compressione del parlato, nel lavoro: McAulay, R.J. and Quatieri T.F., 1986. *"Speech Analysis/Synthesis based on a Sinusoidal Representation"*. IEEE Transactions on Acoustics, Speech and Signal Processing. .

I segnali sonori periodici  - se la loro periodicità è in banda audio - sono suoni *intonati* (dotati cioè di un *pitch)* ed hanno in genere come intonazione la nota che corrisponde alla sua frequenza di ripetizione (un segnale che ha una frequenza di ripetizione di 440 Hz suona ad esempio come un *la* corista). E' opportuno sottolineare che si sta qui introducendo una nozione che non è di tipo fisico-matematico, ma *percettivo*. L'intonazione, o *pitch,* è definita dal nostro sistema auditivo, un sistema complesso e dal funzionamento tutt'altro che chiarito in modo esaustivo. C'è da aspettarsi quindi che la coincidenza tra periodicità e intonazione non sia "matematicamente perfetta", e che si possano incontrare eccezioni e deviazioni dalla norma. Questo il motivo dell'adozione della formulazione prudenziale (*"in genere"*) adottata poco sopra.

Si può affermare che segnali periodici sono intonati, se il loro periodo corrisponde a frequenze in banda audio (20Hz-20KHz). Non è però vero il contrario: segnali che non sono rigorosamente periodici possono suonare come intonati, purché siano in qualche modo "quasi periodici" con "quasi periodicità" nell'ambito audio. Questa seconda nozione, come si vede più oltre, è lasciata volutamente nel vago, perché è impossibile a tutt'oggi  fornire una spiegazione semplice e puramente matematica del fenomeno della percezione dell'intonazione che copra tutti i casi interessanti. Esperienze di percezione dell'intonazione di segnali non periodici si possono fare utilizzando lo *spostamento di frequenza*[3], con il quale è possibile produrre, a partire da segnali perfettamente periodici, segnali non periodici che però continuano ad essere dotati di una intonazione. Questa considerazione diventerà ancora più importante parlando di segnali quasi-periodici.

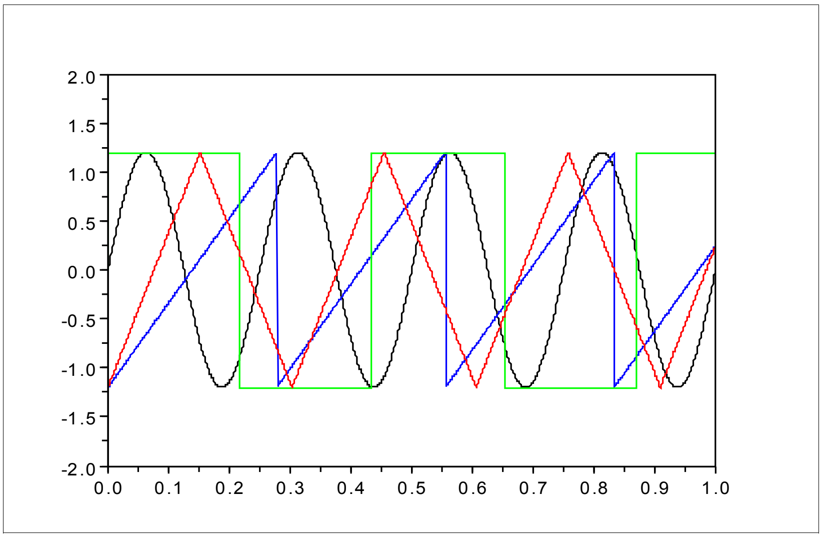


Fig. 1 - Sinusoide (----) , dente di sega (----), quadra (----), triangolo (----)

Esempi di segnali periodici sono le sinusoidi, le onde quadre, il dente di sega, le onde triangolari, i treni di impulsi.

Queste considerazioni spiegano anche perché gli oscillatori tabellari generano sempre suoni intonati, indipendentemente dal contenuto della loro tabella:

3 V. Lorenzo Seno, *Oscillatori*. Dispense, 2007 – Conservatorio de L'Aquila.

perché i suoni generati sono perfettamente periodici, con un periodo pari al tempo di scansione stabilito. Accade così che, caricando dentro un oscillatore siffatto del rumore bianco (o rosa, o marrone), esso generi sul transitorio di attacco un suono di plettro, e una ben precisa nota tenuta.

Mentre con mezzi elettronici è possibile generare segnali perfettamente (matematicamente) periodici, i suoni generati dagli strumenti acustici anche "melodici" lo sono solo approssimativamente, e ciò nonostante essi suonano come "intonati". Siamo qui nel campo della "quasi-periodicità", nel quale è possibile penetrare in questa sede solo in parte.

Partiamo dall'esame dei casi più semplici. Mentre alcuni strumenti (organo a canne, e in misura minore, i fiati o gli archi) sono in grado di produrre suoni "sostenuti[4]" per periodi lunghi o addirittura indefiniti, molti strumenti (a corde pizzicate, come la chitarra o il clavicembalo, o a percussione come il pianoforte o lo xilofono, ad esempio) producono suoni che si estinguono in un intervallo di tempo più o meno breve.

Trascurando temporaneamente altre più sottili deviazioni dalla perfetta periodicità, un suono che "si estingue" può essere definito periodico?

A stretto rigore, applicando la definizione 4, a questa domanda si dovrebbe rispondere di no.

Tuttavia immaginiamo che il processo di estinzione sia *lento,* rispetto al periodo di ripetizione sottostante (intonazione della nota emessa). Dopo un periodo di ripetizione il segnale non sarà esattamente identico al periodo precedente, ma sarà comunque molto simile, si potrebbe dire: identico a meno di un guadagno che nel frattempo, da un periodo all'altro, è variato di poco. Abbiamo qui un primo caso di "quasi-periodicità" dovuta all'ampiezza: un segnale non è perfettamente periodico, ma può essere ricondotto ad una periodicità perfetta compensando un'ampiezza lentamente variabile nel tempo. In queste condizioni non stupisce che il nostro sistema uditivo (così come gli algoritmi di determinazione automatica dell'intonazione) continui a riconoscere l'intonazione "sottostante".

Un altro caso di quasi-periodicità è quello tipico di moltissimi strumenti (archi, fiati, voce) nei quali l'esecutore utilizza il *vibrato.* Il vibrato è una intenzionale, lenta (rispetto alla frequenza d'intonazione) variazione periodica dell'intonazione stessa. Anche in questo caso il segnale non sarà perfettamente periodico, poiché il periodo stesso varia da un periodo all'altro, per così dire. Se però questa variazione è lenta, il periodo resterà quasi costante per un numero elevato di ripetizioni, e noi percepiremo un'intonazione variabile nel tempo, ma "istante per istante" perfettamente identificabile. Il sistema uditivo, per una sorta di principio gestaltico di buona continuazione, "raccorda" questa intonazione variabile nel tempo e la "identifica", separandola dal resto, (*segregate*) come un fenomeno unico variante nel tempo (*time-varian*t).

4 Il termine è qui usato come traduzione del corrisponde inglese: "*sustain*", che indica appunto "costanza di ampiezza". Nella terminologia musicale "sostenuto" è una notazione agogica, e indica in realtà "più lento", l'esatto opposto del significato ordinario, quando si dica ad esempio "a ritmo sostenuto" (cioè, celere).

A causa di inevitabili imperfezioni meccaniche degli strumenti acustici, inoltre, il periodo di ripetizione varierà, sia pure lievemente, da un periodo all'altro, *in modo casuale*. Si tratta dello *jitter di intonazione*, detto talvolta, meno propriamente, *jitter di frequenza*. Se però questo "sbandamento casuale" (*jitter*) è di piccola entità, il sistema uditivo identificherà queste frequenze successive lievemente diverse tra loro *come un'unica frequenza di ripetizione media*, e percepirà il "microsbandamento" come una caratteristica timbrica, dato che la frequenza del fenomeno (la frequenza di variazione dell'intonazione) si trova in banda audio. Nelle canne d'organo, ad esempio, lo sbandamento di intonazione (microfluttuazione) è dovuto nelle canne labiali al *jitter di fase* del getto d'aria, a sua volta dovuto alla turbolenza del flusso d'aria attorno al labbro. Questo dato timbrico distingue un tipo di canna dall'altro a seconda delle particolarità costruttive del labbro, il quale può dare luogo a fenomeni più o meno accentuati di turbolenza a seconda di come sia forgiato il suo bordo.

# 4 Sintesi e analisi di Fourier

Torniamo all'eq. 3, che definisce le funzioni elementari per la sintesi-analisi di Fourier. E' facile convincersi che la sintesi di Fourier produce un segnale risultante *periodico,* con periodo $T=1/\nu_1$ dove $2\pi\nu_1=\omega_1$. Questa circostanza deriva direttamente dalla scelta di utilizzare esclusivamente frequenze in rapporti armonici. Da questa considerazione si potrebbe (erroneamente) dedurre che la sintesi di Fourier è in grado di produrre solo suoni intonati con intonazione fissa. La deduzione è erronea perché le cose possono andare assai diversamente se la periodicità è a frequenze subaudio. Dato che la periodicità di un segnale siffatto non è percepibile come intonazione, possono avvenire in questi casi fenomeni percettivamente e matematicamente complessi.

Ricordiamo qui che effettuare l'analisi di un segnale significa trovare i parametri necessari ad esprimerlo nei termini della corrispondente sintesi. Nel nostro caso, analizzare in termini di Fourier un segnale significa trovare i valori $a_n$ e $\varphi_n$ necessari a rappresentarlo nei termini della 1 (pag. 1), dove le funzioni elementari $s_n(t)$ adottate sono quelle indicate nella 3.

Vale a questo punto un teorema, dovuto a Fourier, di importanza basilare, e che possiamo formulare così:

***(I) Qualunque segnale periodico di periodo T può essere espresso come combinazione lineare di sinusoidi di frequenze armoniche del periodo, eventualmente in numero infinito. Questa combinazione lineare è unica, ovvero ad ogni segnale periodico corrisponde una ed una sola possibilità di analisi in termini di Fourier.***

L'espressione 1 che fa uso di (co)sinusoidi armoniche prende il nome di *serie di Fourier:*

$$\text{output}(t)=\sum_{0}^{\infty} c_n\cdot\cos(n\cdot\omega_1\cdot t+\varphi_n) \qquad 5$$

E' opportuno sottolineare ed esplicitare le numerose implicazioni dell'enunciato I.

Anzitutto esso asserisce che *qualsiasi* segnale periodico è esprimibile in serie di Fourier, cosa che implica la “completezza” di quest'ultima: essa è in grado di esprimere *tutti i segnali periodici*. Questa “completezza” è dovuta a precise proprietà delle funzioni sinusoidali armoniche. In gergo matematico si dice che esse costituiscono una base completa ortonormale per lo spazio di Hilbert rappresentato da tutte le funzioni periodiche di periodo dato. Senza entrare in ulteriori dettagli, il fatto che la base sia ortonormale implica anche l'unicità della espressione: solo una specifica serie esprime il segnale periodico di partenza, ed essa è *l'unica* in grado di farlo. Scegliendo un diverso insieme di frequenze, non armoniche della frequenza di ripetizione del segnale da analizzare, potremmo trovarci nella circostanza di non potere esprimere il segnale di partenza, oppure di poterlo esprimere in più di un modo.

L'enunciato I contiene inoltre la precisazione che può essere necessario un numero infinito di armoniche per esprimere un generico segnale periodico. L'analisi può però essere ristretta ad un numero finito se ci si limita all'ambito dei segnali sonori di interesse musicale. Il nostro sistema uditivo non è infatti in grado di percepire frequenze superiori ad una determinata soglia (convenzionalmente stabilita in 20 KHz). Questo significa che due segnali sonori periodici le cui serie di Fourier siano identiche fino a 20 KHz, ma differiscano per le frequenze superiori, pur essendo matematicamente differenti, saranno percepiti come identici. Tutti i segnali periodici differenti tra loro solo per la zona superiore ai 20 KHz saranno in particolare percepiti come identici a quello tra di loro al quale siano state *soppresse* tutte le frequenze superiori ai 20 KHz, (le loro ampiezze siano state forzate a zero, operazione equivalente a sopprimerle dalla sommatoria). *Nel caso dei segnali sonori non vi è dunque motivo in generale per spingere l'analisi di Fourier al di sopra dei 20 KHz,* e sarà sempre sufficiente uno sviluppo della serie troncato alle frequenze inferiori a questa soglia.

Per effettuare correttamente questo troncamento in sede di analisi, come analizzato nella dispensa relativa ai sistemi campionati, sono però necessarie alcune precauzioni per evitare che le frequenze superiori eventualmente presenti nel segnale da analizzare vadano ad influenzare nel calcolo le ampiezze di quelle inferiori, realizzando così un indesiderato “mescolamento” delle frequenze[5].

Si deve inoltre notare che le funzioni elementari utilizzate, le sinusoidi armoniche, contengono tutte un parametro arbitrario di fase dipendente dall'indice $n$ di armonica. Questo grado di libertà è indispensabile ai fini della completezza, ed è facile convincersene. Se lo si sopprimesse, infatti, avremmo a disposizione solo sinusoidi armoniche con fase zero. Queste valgono tutte zero a $t=0$ e a $t=n \cdot T$, e pertanto anche il segnale risultante dalla loro somma risulterebbe nullo in questi punti. Così facendo, quindi, la serie sarebbe in grado di esprimere solo quei segnali periodici di periodo *T* che valgono zero al tempo zero, (e conseguentemente, zero al tempo *T*), mentre sarebbe impossibile esprimere segnali sempre di periodo *T* ma con un valore arbitrario al tempo zero.

5 Si tratta del fenomeno del *fold-over*, o *aliasing*, o ancora detto delle *frequenze fantasma*.

## 4.1 Componente continua

Una speciale menzione merita nella 3 il termine con $n=0$. Esso si riduce a $s_0(t)=\sin(\varphi_0)$, cioè ad un *termine costante,* non dipendente dal tempo. Questo termine rende il segnale non simmetrico attorno allo zero, e prende anche il nome di *componente continua* (termine mutuato dal gergo elettronico). Esso è pari alla media del segnale nell'intervallo di tempo di analisi o sintesi definito dalla fondamentale. *Segnali privi di componente continua sono dunque a media nulla.* L'eventuale presenza di questo termine (che deve essere moltiplicato come tutti gli altri per il corrispondente coefficiente, dando luogo ad un valore arbitrario) ha come effetto quello di "alzare" o "abbassare" (a seconda del segno di $\sin(\varphi_0)$ ) il segnale restante, e *non è udibile*, poiché noi percepiamo come suono solo le *variazioni* nel tempo (con frequenze più veloci di 20 Hz, convenzionalmente) della pressione atmosferica, non il suo valore statico. Per questo motivo questa componente viene spesso omessa nei testi di elaborazione del segnale musicale. Può essere però pericoloso dimenticarsene, perché determinati processi di elaborazione possono generare componenti continue anche a partire da segnali che non ne hanno[6]. E' quindi bene ricordarsi sempre che *in linea di principio la componente continua* (o componente costante, a frequenza zero, o valore medio) fa sempre parte del "pettine" di frequenze armoniche della sintesi-analisi di Fourier.

## 4.2 Formulazioni polare e cartesiana della serie di Fourier.

Effettuare quindi un'analisi di Fourier di un segnale periodico con periodo *T* significa dunque determinare *due parametri* per ogni frequenza armonica: l'ampiezza $a_n$ e la fase $\varphi_n$. Esiste a questo punto una formulazione alternativa (ma del tutto equivalente) della serie di Fourier che fa uso di una funzione elementare più "simmetrica". Abbiamo quindi il seguente (importante) enunciato:

***(II) Una sinusoide (oppure una cosinusoide) con ampiezza e fase può essere a sua volta espressa come combinazione lineare di una sinusoide e di una cosinusoide ciascuna con fase 0:***

$$c\cdot\sin(\phi+\varphi)=a\cdot\sin(\phi)+b\cos(\phi) \qquad 6$$

In altre parole, possiamo esprimere il nostro segnale periodico nei termini di una serie di sinusoidi e cosinusoidi entrambe con fase zero e ciascuna con una sua propria ampiezza. Questa seconda formulazione è del tutto equivalente, e come ovvio non cambia il numero di parametri da determinare nell'analisi: sempre due per ogni frequenza armonica in gioco.

$$output(t)=\sum_{1}^{\infty}(a_n\cdot\sin(n\cdot\omega_0\cdot t)+b_n\cdot\cos(n\cdot\omega_0\cdot t)) \qquad 7$$

La rappresentazione 5 è detta *polare,* la 7 è detta *cartesiana.*

Il passaggio dall'una all'altra rappresentazione è facilmente ottenibile ricordando le formule di addizione di trigonometria:

6 V. Lorenzo Seno, *La modulazione di ampiezza*. Dispense (2006).

$\sin(\phi+\varphi)=\sin(\phi)\cdot\cos(\varphi)+\cos(\phi)\cdot\sin(\varphi)$

Otteniamo dunque la 6 ponendo:

$a=c\cdot\cos(\varphi)$ e $b=c\cdot\sin(\varphi)$

e riscrivendo la 6 in questo modo:

$$output(t)=\sum_{n=0}^{N} c_n\cdot(\cos(\varphi_n)\cdot\sin(n\cdot\omega_1\cdot t)+\sin(\varphi_n)\cdot\cos(n\cdot\omega_1\cdot t))$$

Se esaminiamo ora il termine generico della 5 ( $c_n\cdot\sin(n\cdot\omega_1\cdot t+\varphi_n)$ ), osserviamo che questo può essere interpretato come una *rotazione* di un angolo $\varphi_n$ del fasore $e^{in\cdot\omega_1 t}$ e della applicazione di un guadagno $c_n$. Questo termine generico può quindi essere considerato come la parte immaginaria del prodotto del fasore $e^{in\cdot\omega_1 t}$ per il fasore $c_n\cdot e^{i\varphi_n}$:

$$c_n\cdot\sin(n\cdot\omega_1\cdot t+\varphi_n)=\Im(e^{in\cdot\omega_1 t}\cdot c_n e^{i\varphi_n})$$

Questo spiega l'usanza (ormai consolidata) di esprimere le trasformate di Fourier nel campo complesso, con il termine $c_n e^{i\varphi_n}$ ma in modo cartesiano:

$$c_n e^{i\varphi_n}=c_n\cdot(\cos(\varphi_n)+i\sin(\varphi_n))$$

e fornendo la coppia:

$\Re_n=c_n\cos(\varphi_n)$ $\Im_n=c_n\sin(\varphi_n)$

ovvero il numero complesso: $\Re_n+i\Im_n$

Se si vuole però conoscere l'ampiezza della componente *n*-esima (il termine *n*-esimo dello *spettro)*, va calcolato $c_n$. Nei sistemi che presentano le trasformate di Fourier come numeri complessi, si tratta di eseguire il modulo:

$$c_n=|\Re_n+i\Im_n|$$

Nei sistemi non dotati di notazione complessa la trasformata viene fornita come coppia di valori reali ( $\Re_n$, $\Im_n$ ), e per ottenere $c_n$ va quindi eseguito il calcolo dell'ipotenusa, utilizzando il teorema di Pitagora. Questa operazione viene detta in gergo *conversione cartesiano-polare,* che in genere è presente come funzione elementare nei sistemi numerici di elaborazione del segnale, la quale è munita di due ingressi e fornisce due uscite: il modulo e la fase.

In mancanza della relativa funzione il calcolo è peraltro semplice:

$$c_n=\sqrt{\Re_n^2+\Im_n^2} \quad 8$$

mentre per la fase si ha:

$$\varphi_n=\arctan\left(\frac{\Im_n}{\Re_n}\right) \quad 9$$

Nei sistemi di elaborazione di segnale esiste generalmente anche la funzione inversa, la *conversione polare-cartesiano,* la quale effettua il calcolo opposto da modulo e fase a componenti reale-immaginario:

$\Re_N = c_n \cdot \cos(\varphi_n) \qquad \Im_n = c_n \cdot \sin(\varphi_n)$

Le operazioni in 8 e 9, che sono implicite anche nelle funzioni cartesiano-polare, sono piuttosto dispendiose in termini di calcolo. E' bene tenerlo presente per usarle con la dovuta parsimonia.

La 9 comporta inoltre una difficoltà intrinseca, anche se solo apparente: quando $\Re_n = 0$ l'argomento della funzione arcotangente diventa infinito. La difficoltà è solo apparente perché questa circostanza si verifica quando $\varphi_n = n\pi$ , e dunque il valore della funzione è noto senza bisogno di passare per il calcolo dell'argomento. Questo ha portato talvolta all'introduzione di una funzione "adattata" per risolvere il caso $\Re_n = 0$ , munita di due argomenti separati per $\Im_n$ e $\Re_n$ , talvolta chiamata "atan2". Questa funzione gestisce in modo opportuno l'intorno di $\Re_n = 0$ per fornire un valore corretto dell'argomento senza passare per il calcolo del rapporto, che condurrebbe a valori enormi (al limite infiniti), dell'argomento. Le funzioni di conversione cartesiano-polare di norma gestiscono correttamente questa circostanza.

# 5 Trasformata e spettro.

L'operazione di analisi (calcolo dei coefficienti della serie di Fourier) prende anche il nome di *Trasformata di Fourier*. L'operazione di sintesi prende anche il nome di *antitrasformata (di Fourier)*. Dato che la serie di Fourier che così si ottiene è del tutto equivalente al segnale originale (si passa in modo del tutto reversibile dall'uno all'altra mediante trasformazione o antitrasformazione), e i coefficienti che così si ottengono rappresentano le ampiezze delle componenti in frequenza (armoniche), si dice anche che la trasformazione rappresenta il segnale *nel dominio della frequenza,* mentre l'antitrasformazione lo fa *nel dominio del tempo*. In termini della rappresentazione polare 5, il grafico delle ampiezze $c_n$ in funzione della frequenza $\omega_n$ si chiama *spettro* del segnale. Il grafico di $c_n^2$ prende il nome di *spettro di potenza* del segnale. Va fatto notare che la sola conoscenza dello spettro o dello spettro di potenza non è sufficiente alla ricostruzione del segnale originale: è infatti necessario conoscere anche l'informazione di fase $\varphi_n$ . Sotto certe condizioni però (quando la periodicità sia sufficientemente breve, inferiore ai 20 msec), nel caso di segnali sonori destinati all'ascolto umano, questa informazione può essere omessa (posta artificiosamente a zero), perché il nostro sistema uditivo *non è sensibile alla fase* su queste scale di tempo. In altre parole, in queste condizioni, segnali con uguale spettro ma con fasi differenti sono matematicamente diversi, ma vengono percepiti come uguali.

Una breve digressione: questa è una delle basi sulle quali si fonda la compressione Mp3 del segnale audio, la quale opera secondo queste linee generali. Il segnale viene trasformato e memorizzato (o trasmesso) nel dominio della frequenza. Si realizza un risparmio di informazioni attraverso l'omissione dell'informazione di fase, l'uso di metodi di codifica ottimi che ripartiscono i bit a disposizione in modo da utilizzarne molti per le informazioni spettrali rilevanti, e pochi per quelle di dettaglio, e infine la soppressione di informazioni spettrali non percepibili o scarsamente percepibili sulla base di fenomeni psicoacustici come gli effetti di mascheramento.

Mentre l'operazione di sintesi è facilmente concepibile, ci si può a questo punto domandare, dato un segnale periodico arbitrario (di periodo *noto* $T=2\pi/\omega_1$ ), come si possa eseguire l'analisi di Fourier, ovvero calcolare i coefficienti $a_n$ e $b_n$; in poche parole, come calcolare la Trasformata di Fourier.

Senza dilungarci in dimostrazioni, supponendo di volere analizzare il segnale $\text{output}(t)$, il calcolo si riduce ad un paio di integrali definiti:

$$a_n=\frac{1}{\pi}\int_{\frac{-\pi}{2\omega_1}}^{\frac{\pi}{2\omega_1}} \text{output}(t)\cdot\sin(n\cdot\omega_1\cdot t)\,dt \quad a_n=\frac{1}{\pi}\int_{-\frac{T}{2}}^{\frac{T}{2}} \text{output}(t)\cdot\sin(n\cdot\omega_1\cdot t)\,dt \tag{10}$$

$$b_n=\frac{1}{\pi}\int_{\frac{-\pi}{2\omega_1}}^{\frac{\pi}{2\omega_1}} \text{output}(t)\cdot\cos(n\cdot\omega_1\cdot t)\,dt \quad \mathrm{b}_n=\frac{1}{\pi}\int_{-\frac{T}{2}}^{\frac{T}{2}} output(t)\cdot\cos(n\cdot\omega_1\cdot t)\,dt \tag{11}$$

In altre parole, ciascun coefficiente è pari all'integrale del prodotto del segnale da trasformare per la relativa funzione di base, *in un periodo* della fondamentale (dell'intervallo temporale di analisi) centrato attorno all'origine dei tempi.

Ovviamente, per quanto sopra detto, ci si può limitare al calcolo dei soli termini per i quali $n\cdot\omega_1<2\pi\cdot 20\,KHz$ .

Il valore delle 10 e 11 è puramente teorico, se $\text{output}(t)$ non è una funzione matematica espressa in forma chiusa. Se si tratta di un suono *concreto,* al giorno d'oggi esso sarà espresso in forma numerica[7], dunque come *sequenza* di valori. Il calcolo degli integrali 10 e 11 dovrà essere quindi eseguito in forma numerica (*DFT, Discrete Fourier Transform*), e questa circostanza introduce nuovi problemi di natura sia teorica, sia pratica, i primi legati al teorema del campionamento, e i secondi al numero (elevato) di calcoli necessari.

Il teorema del campionamento lega in modo indissolubile il numero di campioni contenuti nel periodo *T* al numero di armoniche calcolabili: le armoniche sono pari a $N/2$ se ***N*** è il numero di campioni. Si tratta pertanto di calcolare ***N*** integrali definiti in modo numerico, e anche adottando un algoritmo molto semplice per quest'ultima incombenza (suddividendo l'integrale in N punti) il numero di calcoli da eseguire, come si vedrà meglio nel seguito, cresce come $N^2$ .

Fortunatamente un algoritmo particolarmente efficace (*FFT, Fast Fourier Transform*), dovuto all'astuzia di Gauss e alla sua riscoperta all'epoca dei calcolatori elettronici fatta da Cooley e Tukey, permette di ridurre il numero di calcoli necessari in modo tale che essi diventino proporzionali a $N\cdot\log_2(N)$ , il che è drasticamente meglio di $N^2$ , per ***N*** grande. Questa serie di problemi sarà affrontato nel capitolo apposito. Per ora daremo per scontato (trattandosi in effetti in questo contesto di una *technicality*) che il calcolo della trasformata di Fourier

7 Anche un segnale di sintesi sarà espresso presumibilmente in forma numerica, a meno che non provenga da un sintetizzatore analogico. In tal caso esso sarebbe, ai fini del nostro discorso, del tutto equivalente ad un suono concreto.

non rappresenti un problema[8].

## 5.1 *Relazioni tra spettro e forme d'onda*

Abbiamo visto a titolo di esempio nel cap. 3 come diversi andamenti dello spettro e della fase corrispondano a diverse forme d'onda, e viceversa.

La determinazione esatta di quale spettro corrisponda ad una determinata forma d'onda, e viceversa, può ovviamente essere fatta solo eseguendo l'operazione di trasformazione o di antitrasformazione. Esistono però alcune relazioni qualitative che permettono di stabilire alcune caratteristiche spettrali a partire da un esame sommario della forma d'onda. Il contrario è invece assai più difficile.

Ricordiamo che lo spettro (ottenuto dalla trasformata di Fourier) esprime in funzione della frequenza l'ampiezza del contributo delle diverse sinusoidi armoniche nell'operazione di ricostruzione del segnale originale mediante combinazione lineare (somma pesata) delle stesse.

I segnali elementari che costituiscono la nostra base sono funzioni che hanno determinate caratteristiche: in particolare, sono continue (cioè prive di salti) e prive di spigoli vivi. Data una determinata frequenza, la sinusoide di ampiezza unitaria ha una sua massima "ripidezza" (massima derivata, corrispondente al punto di attraversamento dello zero) e una sua massima curvatura (corrispondente ai punti di massimo e di minimo).

Una discontinuità è una salto istantaneo da un valore ad un altro, e quindi dà luogo ad una pendenza infinita in quel punto. Un punto angoloso (spigolo vivo) è invece un punto dove la curvatura è infinita. Se la forma d'onda originale contiene delle discontinuità o dei punti angolosi, viene spontaneo domandarsi come si possa riuscire a riprodurre comportamenti del genere sommando funzioni, quali le sinusoidi, che sono del tutto prive di caratteristiche del genere.

Esiste però una sinusoide dotata sia di pendenza, sia di curvatura, infinite: è la sinusoide a frequenza infinita. Basta immaginare il processo al limite: più è alta la frequenza, più la pendenza (nei punti di attraversamento dello zero) si avvicina alla verticale (pendenza infinita), e più la curvatura nei punti di massimo e minimo si avvicina all'infinito (spigolo vivo).

Per riprodurre discontinuità o punti angolosi è quindi necessario "spingersi" fino a frequenza infinita, ovvero utilizzare un numero infinito di armoniche. Questo conduce anche alla "ricetta" qualitativa secondo la quale il contributo delle frequenze alte sarà tanto maggiore quanto più nel segnale sono presenti pendenze ripide o curvature alte.

In un suono quella caratteristica percettiva che dipende dal rapporto tra ampiezza delle frequenze alte e quelle basse prende il nome di *"sharpness"*. Queste considerazioni ci permettono di stimare qualitativamente "ad occhio" la *sharpness* esaminando la forma d'onda.

---

8 In effetti, esiste una pluralità di sistemi software "in scatola", anche *Open Source*, per eseguire il calcolo della trasformata e antitrasformata di Fourier, sia ad alto livello (pacchetti per il suono o il calcolo muniti di GUI), sia a basso livello (funzioni di libreria).

Inoltre, un segnale il cui spettro si estenda fino all'infinito è detto *a banda illimitata (e a banda limitata* in caso contrario). Un segnale del genere, come noto, non è rappresentabile in un sistema campionato nel tempo. Queste considerazioni ci permettono quindi di stimare se un segnale sia a banda illimitata: la presenza di pendenze "infinite" e di punti angolosi indicano inequivocabilmente che il segnale è a banda illimitata.

Negli esempi di pag. 3, ad esempio, sia l'onda quadra che il dente di sega sono evidentemente a banda illimitata, trattandosi di segnali dotati di discontinuità. Anche il triangolo è a banda illimitata perché è dotato di punti angolosi.

Maggiore prudenza deve però essere usata nella stima opposta: l'assenza di singolarità come quelle qui descritte non è sicuro indice che il segnale sia a banda limitata.

Un segnale ad esempio ottenuto dalla sequenza di quarti di circonferenza superiori e inferiori non ha né discontinuità né punti angolosi, e non ha nemmeno pendenze infinite, ma è anche esso a banda illimitata[9].

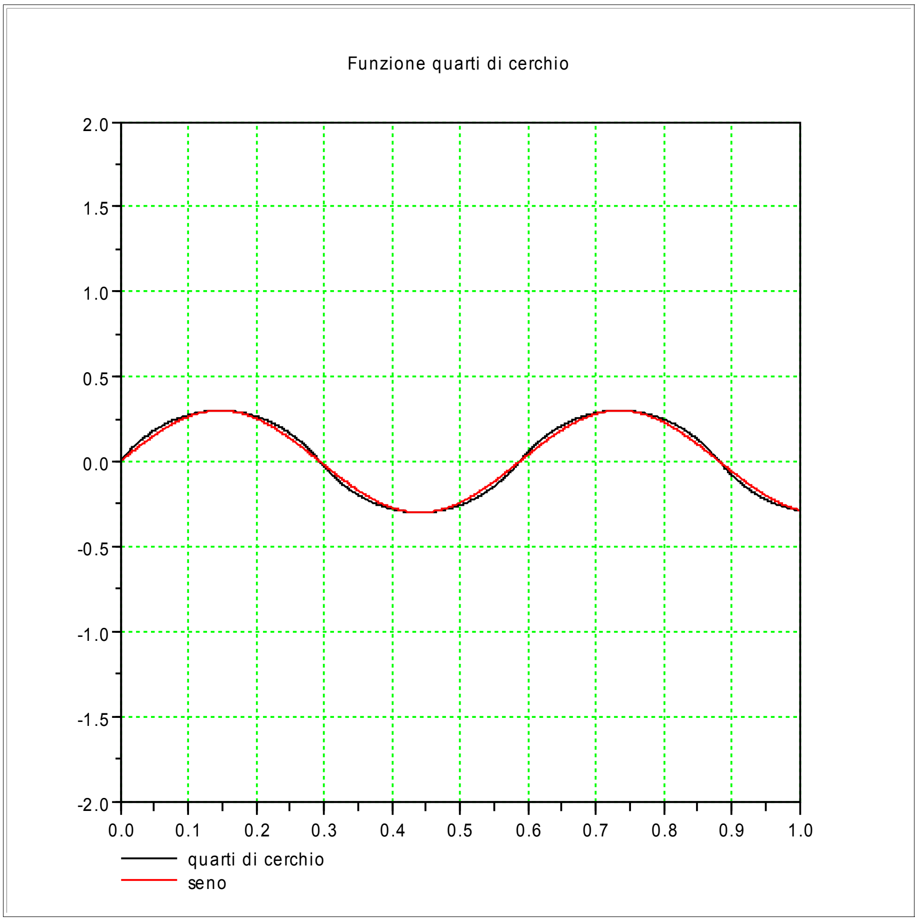


Fig. 2 - Quarti di cerchio e seni

9 Qualcuno potrebbe però notare che è dotato di discontinuità nelle derivate seconde (nell'attraversamento dello zero), altra caratteristica assente nelle sinusoidi, che non presentano discontinuità nelle derivate di ogni ordine.

### 5.2 *Periodicità e intonazione nella rappresentazione di Fourier*

Torniamo ora alle considerazioni sui rapporti tra periodicità e intonazione. A un primo approccio alla sintesi di Fourier, come già detto, molti sono indotti a pensare che, a causa dell'armonicità delle componenti utilizzate e della conseguente periodicità del segnale generato (periodicità *matematica*, esatta), la sintesi di Fourier, se utilizzata in modo tempo-invariante, generi necessariamente dei "tappeti" sonori uniformi, privi di qualunque forma di prosodia, magari necessariamente intonati. Le cose non stanno affatto così, e per convincersene basta rivolgersi per un attimo all'analisi.

Anzitutto costruiamo un segnale artificiale, perfettamente periodico, quello che in gergo si chiama uno "*sweep*" (o "*chirp*"): un segnale sinusoidale che parta ad una determinata frequenza, supponiamo 20 Hz, e la cui frequenza cresca linearmente con il tempo fino a 20 KHz nel giro di un secondo esatto, per poi ripetersi ciclicamente all'infinito.

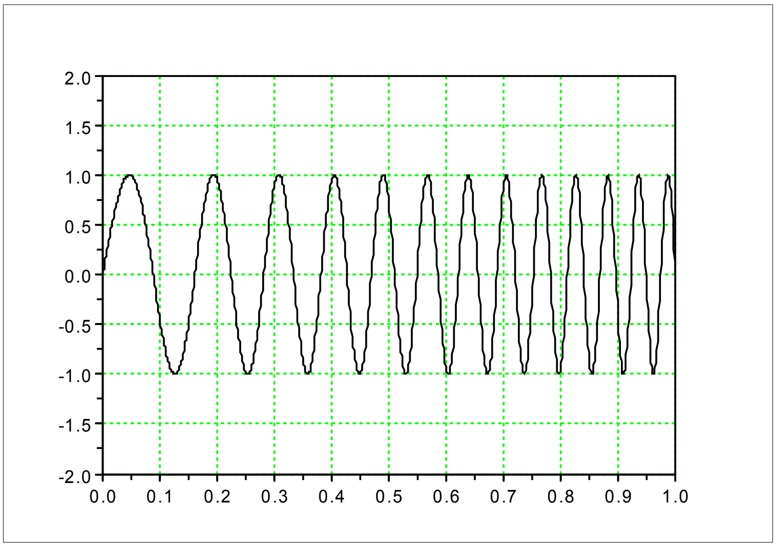


Fig. 3 - Esempio di *sweep* da 5 Hz a 20 Hz in 1 secondo.

Segnali del genere hanno diversi usi pratici: sono utilizzati ad esempio come stimolo per caratterizzare una sala dal punto di vista della risposta acustica, perché "esplorano" tutte le frequenze. A causa della particolare facilità con il quale un segnale del genere si presta alla determinazione del tempo di ritardo con il suo eco, esso è utilizzato sia negli "apparati sonar" degli animali che ne fanno uso (delfini, pipistrelli), sia dei sonar artificiali (come quelli marini), sia infine dei radar.

Ora un segnale come quello che abbiamo appena costruito è perfettamente periodico, con periodo di 1 secondo, pari ad una frequenza di ripetizione di 1 Hz. Secondo l'enunciato I di pag. 5, esso è trasformabile in termini di frequenze armoniche di 1 Hz: 1, 2, 3, 4 ... 20.000 Hz.

Supponiamo ora di risintetizzare questo segnale procedendo alla somma di queste (20.000) sinusoidi: otterremo esattamente, matematicamente, il segnale "chirpato" originale. Sommando cioè in modo opportuno, ma comunque *tem-*

*po-invariante*, il contributo di questi oscillatori armonici ad ampiezza e frequenza fissa otterremo un segnale con intonazione variabile nel tempo e con una sua prosodia (sia pure, in questo specifico caso, elementare).

Non è necessario però limitarsi a considerare un segnale così elementare: possiamo pensare ad una nota dotata di un forte vibrato, di tremolo e messa di voce della durata di un secondo, ad esempio. Oppure anche di un colpo di tamburo (un suono, cioè, senza intonazione). Facendone la trasformata e poi risintetizzandolo, otterremmo un suono periodico (con periodo di un secondo) il quale ripercorrerebbe, all'interno di ciascun secondo, la prosodia originale, e dunque riprodurrebbe – nel caso specifico - a distanza di un secondo ripetutamente il colpo di tamburo.

Ovviamente in questi casi, essendo la base temporale (periodo di ripetizione) piuttosto lungo, molto più lungo dei 20 msec., l'informazione di fase è essenziale: se alterassimo nella sintesi le fasi ottenute dall'analisi, ad esempio nel nostro primo caso non otterremmo il medesimo *sweep* di partenza, ma un segnale con un'altra prosodia. Al limite, otterremmo qualcosa di simile ad un impulso con periodo di un secondo: sia lo *sweep*, sia l'impulso, hanno infatti spettri molto simili, piatti, ma ciò che li differenzia sono fondamentalmente le relazioni di fase tra le armoniche.

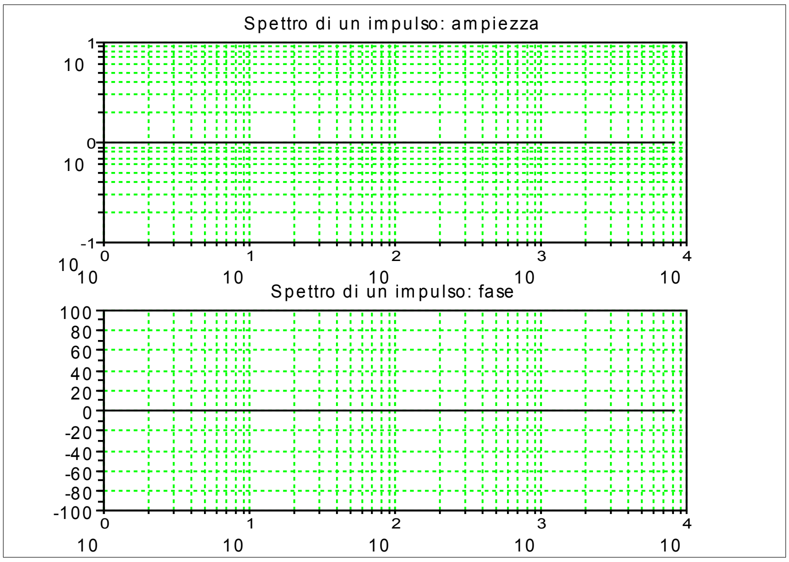


Fig. 4 - Spettro di un impulso di ampiezza 1

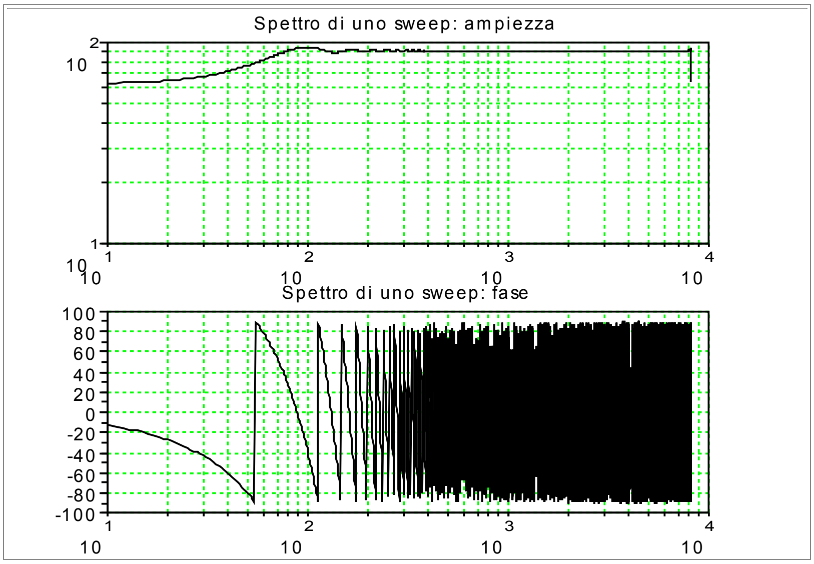


Fig. 5 - Spettro di uno sweep di ampiezza 1, da 0 a 8 KHz.

A proposito delle due figure qui sopra si devono notare alcuni aspetti importanti:

1. Lo *sweep* (a differenza dell'impulso) non ha componente continua, a 0 Hz l'ampiezza dello spettro è zero: l'ampiezza accenna infatti una discesa a zero verso sinistra.
2. La fase dello *sweep* (a differenza di quella dell'impulso che è costante e uguale a zero)  varia molto rapidamente. La rappresentazione utilizzata (la funzione arcotangente) "confina" la fase tra -90° e +90°, e quindi produce quei "salti" che sono in realtà "artefatti matematici". Si deve pensare alla fase come "raccordata" eliminando quei salti (il punto a destra, del salto, a +90°, coincide con quello a sinistra a -90°).
3. L'ampiezza dello spettro dello *sweep* è circa 100 volte quella dell'impulso, anche se lo *sweep* utilizzato ha ampiezza uno esattamente come l'impulso. Questo è il motivo per il quale si preferisce lo *sweep* come segnale di eccitazione per misurare il comportamento in frequenza di sistemi (dispositivi elettronici, sale da concerto, casse acustiche, ecc.): lo *sweep* è un segnale molto più potente, a parità di massima ampiezza disponibile, e quindi fornisce un (molto) migliore rapporto segnale/rumore.

Non è quindi in generale vero che sommando oscillatori armonici "fissi" si ottiene un "tappeto" omogeneo. Se si riflette un attimo al meccanismo di sommazione, ci si convince facilmente che le relazioni tra le fasi delle armoniche intervengono nella prosodia, dato che sono le fasi a governare la coincidenza tra i picchi o gli zeri delle diverse sinusoidi (si tratta del fenomeno di *interferenza)*. A questo "cooperano" le ampiezze: infatti due componenti in frequenza si "indeboliscono" reciprocamente nelle zone temporali nella quali hanno fase opposta (in *controfase*); possono invece rafforzarsi se hanno fasi "dallo stesso lato" (sopra, sotto, se si tratta di sinusoidi) del cerchio goniometrico.

## 6 Analisi di Fourier di segnali dei quali non è nota la periodicità, oppure non periodici: la STFT (*Short Time Fourier Transform*)

Nelle considerazioni svolte nel capitolo precedente, e nei relativi calcoli, la periodicità del segnale da analizzare era data come nota. Ci si può domandare ora come procedere quando questa periodicità non sia nota, oppure si abbia a che fare con segnali quasi-periodici oppure non periodici.

Se si considerano le analisi *in tempo reale*, non effettuate cioè su segnali registrati (e quindi noti al momento della analisi nella loro interezza), è in effetti raro che si possa effettuare un'analisi "sincrona all'intonazione" (*pitch-synchronous*), perché la genesi del segnale è ignota e a priori indeterminata, o la periodicità del segnale potrebbe essere variabile nel tempo, oppure ancora questo potrebbe essere non periodico, o costituito dalla somma di segnali a diversa periodicità (come nel caso degli *accordi*), ovvero con *periodicità diverse sovrapposte*.

Se si analizza in termini di Fourier un segnale periodico su di un intervallo temporale che non è uguale alla periodicità, ci si trova nella circostanza di esprimere una sequenza armonica di periodo $T$ nei termini di una sequenza armonica differente, di periodo $T'$. Ciò è sempre possibile, perché è sempre possibile "periodicizzare" il segnale di periodo $T$ "troncandolo" sulla periodicità $T'$, ed è esattamente quello che si fa quando si calcola la trasformata assumendo come base $T'$. Si otterrà un segnale periodico di periodo $T'$ ottenuto dalla ripetizione della frazione del segnale originario compresa tra $0$ e $T'$. Tuttavia lo spettro del segnale così riperiodicizzato – definito nei termini delle armoniche di $T'$, e non di $T$ – mantiene una parentela piuttosto stretta con lo spettro "originale", ed è quindi in grado di informarci comunque sulle caratteristiche spettrali "originarie". Lo spettro a base $T'$ presenterà infatti una sequenza di picchi e avvallamenti che "tentano" di riprodurre lo spettro originale. Il fenomeno è tanto più decifrabile quanto più $T'$ è maggiore di $T$, ovvero quanto più bassa è la frequenza fondamentale d'analisi.

Supponiamo ad esempio di analizzare un segnale con una base temporale di 0.1 sec (corrispondente ad una fondamentale di analisi di 10 Hz), e supponiamo ora che questo segnale sia composto di una sola sinusoide a frequenza 100 Hz. Ci troviamo nella fortunata situazione in cui la frequenza da descrivere è esattamente una della armoniche della fondamentale d'analisi (la decima armonica). Nello spettro ottenuto per trasformazione troveremo tutti i valori a zero, salvo quello relativo alla decima armonica. Si tratta dell'ovvia circostanza che per produrre una frequenza a 100 Hz è necessario un solo oscillatore a 100 Hz.

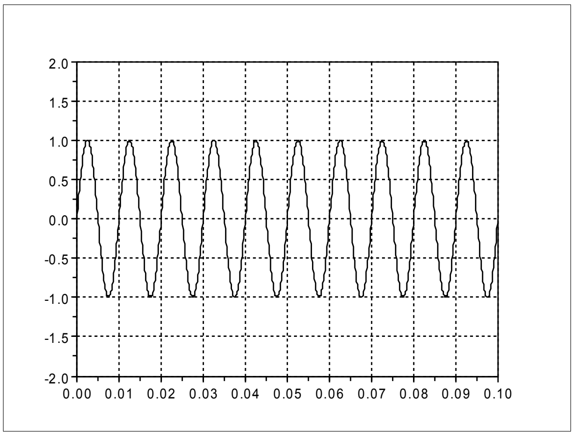


Fig. 6 - Sinusoide a 100 Hz, della durata di 0.1 sec.

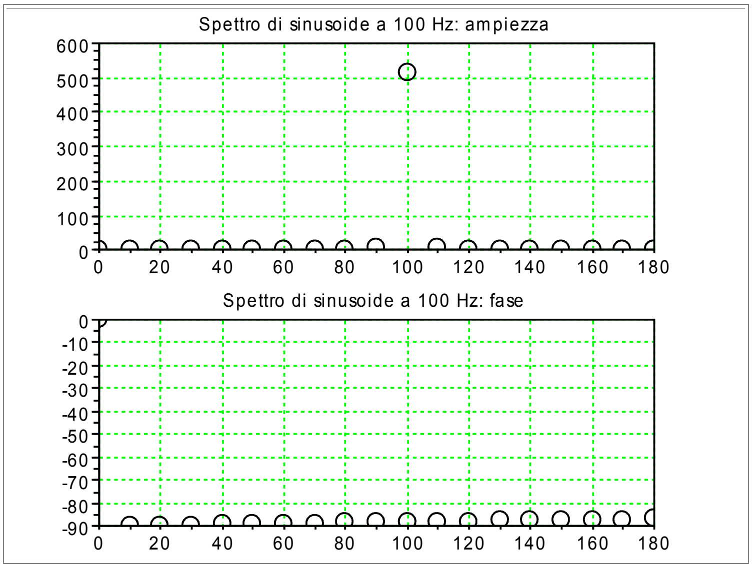


Fig. 7 - Dettaglio della trasformata polare di Fourier di una sinusoide a 100 Hz su 0.1 sec..
(il grafico utilizza punti, dato che si tratta di un insieme discreto di frequenze)

Supponiamo ora invece che il segnale da analizzare sia costituito da una frequenza di 104 Hz. Tale frequenza non è presente nelle armoniche di analisi, ma ve ne sono due molto vicine: 100 Hz e 110 Hz. Nello spettro ottenuto per trasformazione sulla base di 0.1 sec. troveremo particolarmente accentuate proprio queste due frequenze "vicine", che con il loro "addensamento" indicano che nel segnale originale c'è una notevole presenza di segnale nei dintorni di quei valori. Troveremo che anche tutte le altre frequenze (fino alla 10 Hz e la continua a 0 Hz) hanno un'ampiezza più piccola ma diversa da zero, perché la riproduzione del segnale con una frequenza fuori dal pettine "mobilita" tutto lo spettro a disposizione.

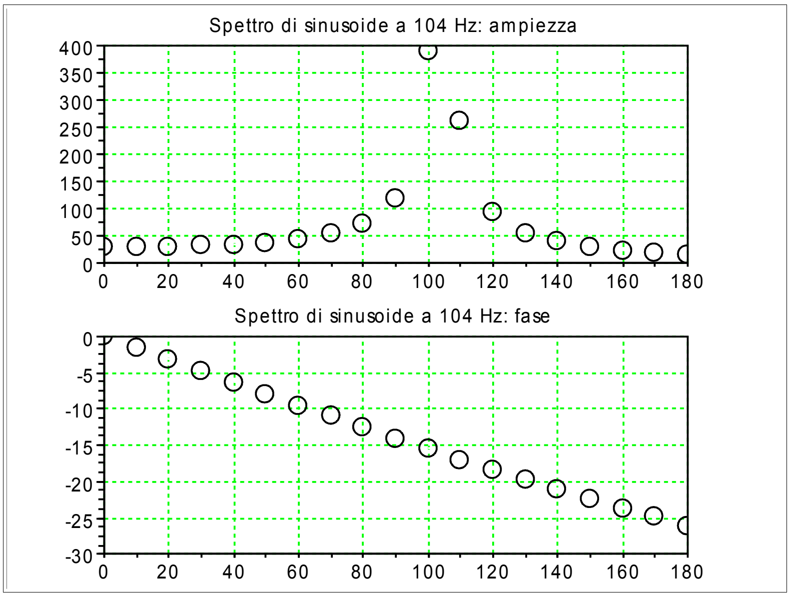


Fig. 8 - Dettaglio della trasformata polare di Fourier di una sinusoide a 104 Hz su 0.1 sec.

La presenza della continua (che nessuno ha inserito nel segnale originale) non deve stupire: un segnale sinusoidale a 104 Hz su di una base di 0.1 sec non inizia e termina con la stessa fase: inizia con fase 0, ma termina con una fase diversa da zero. Su quella base temporale *non è quindi simmetrico attorno allo zero,* e non ha dunque media nulla: ha una componente continua dovuta proprio a questo effetto di “troncamento”. Il segnale “visto” da questa finestra inoltre possiede un “salto” brusco alla fine. Per riprodurre questo brusco salto sono necessarie le frequenze “più alte”, teoricamente fino all'infinito se pensiamo il fenomeno nel continuo[10].

10 Nel discreto, o nell'ambito numerico, le frequenze disponibili sono invece sempre in numero finito, per il teorema di Nyquist. Se il segnale è campionato temporalmente, i salti iniziali o finali non potranno essere “infinitamente ripidi” dato che come minimo implicheranno una distanza temporale di un periodo di campionamento.

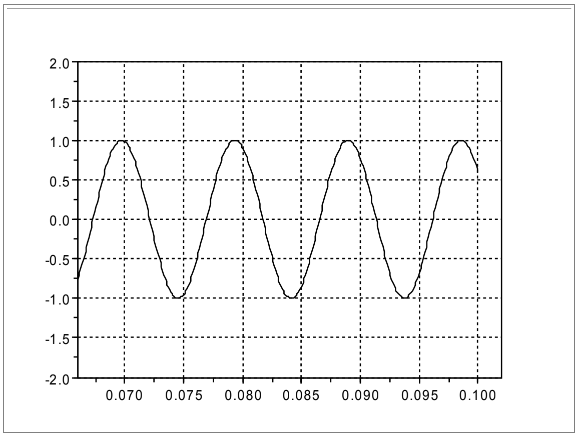


Fig. 9 - Dettaglio finale di una sinusoide a 104 Hz. su di un intervallo di 0.1 sec.

Un frammento temporale di durata *T'* di un segnale qualunque può essere considerato il "rappresentante" di un periodo di un segnale periodico con periodo *T'* ottenuto replicando infinite volte a destra e a sinistra quel frammento. La trasformata di Fourier di questo ipotetico segnale periodico ci fornisce comunque, in base alle considerazioni fatte più sopra, delle informazioni sulla composizione "sottostante" del segnale, fino a permetterci in linea teorica di ricostruire la periodicità "vera" del segnale originario (se questo è periodico). E' però piuttosto intuitivo che perché questa cosa sia possibile, nell'intervallo temporale di analisi devono essere presenti in linea strettamente teorica almeno due dei periodi originari[11]. In pratica, il compito diventa tanto più fattibile quanto più il periodo di analisi è superiore alla periodicità originaria.

Queste considerazioni forniscono la base per la tecnica di analisi detta STFT (*Short Time Fourier Transform* – Trasformata di Fourier su breve periodo). Ha in pratica senso su di un segnale qualunque eseguire una trasformata di Fourier su di una base temporale arbitraria, per segmenti successivi (ovvero, come si dice, per *frame).* Otterremo quindi una successione temporale di spettri, quello che viene chiamato *spettrogramma*. Abbiamo così una rappresentazione *mista:* sia nel tempo, sia nella frequenza. Per rappresentare questa successione di funzioni della frequenza sistemate su di un asse temporale, si preferisce spesso rappresentarle anziché con una successione di grafici, con una successione di strisce grigie o colorate, adottando una scala di intensità di grigio o di "falso-colore". Si ottiene così un notevole colpo d'occhio: le frequenze "sottostanti" nel segnale sono evidenti come zone di colore chiaro (o più scure, quando si adotti una rappresentazione a scala di grigi), permettendo di individuare "ad occhio" le frequenze costituenti, anche quando siano "fuori passo" rispetto alla finestra temporale di analisi.

11 E' ovviamente impossibile riconoscere una periodicità se non si esaminano almeno due periodi completi.

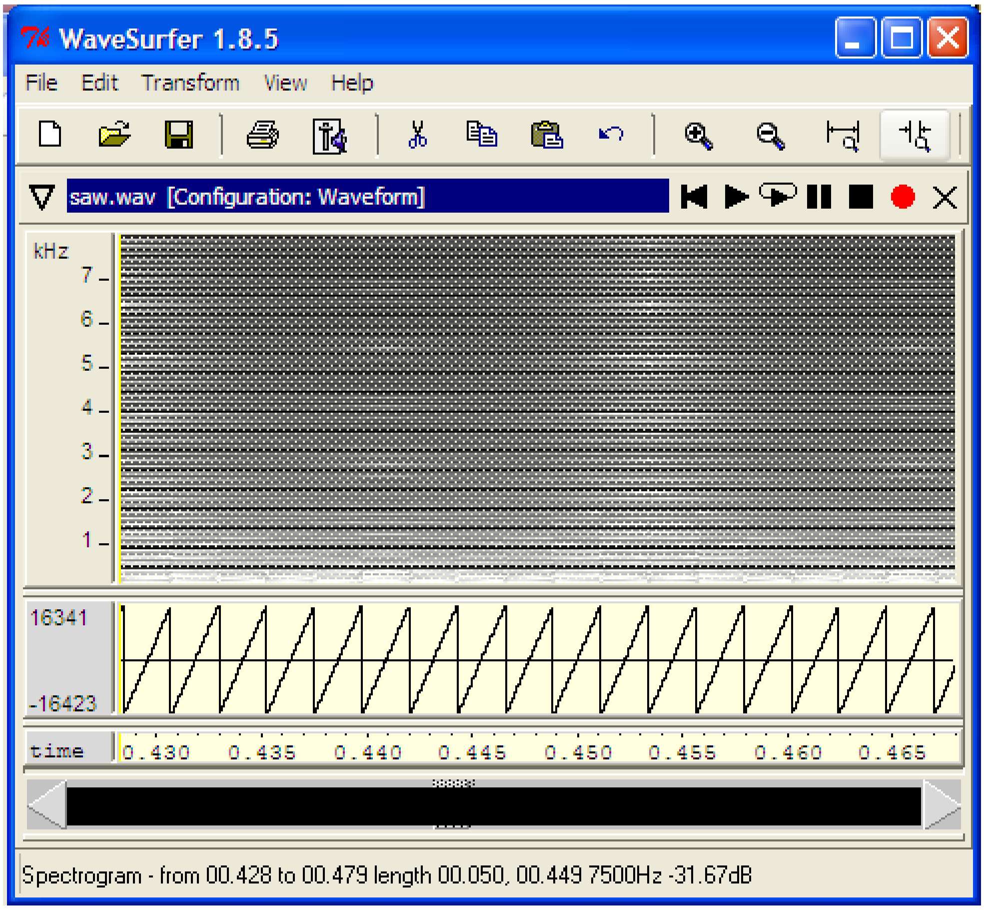


Fig. 10 - WaveSurfer mostra lo spettrogramma di un dente di sega a 440 Hz.

## *6.1 Relazioni tempo-frequenza*

Abbiamo visto che l'intervallo di analisi *T* determina la fondamentale di analisi, e quindi il passo del pettine di armoniche utilizzato. L'analisi di Fourier fa uso di un insieme discreto di frequenze, tra le quali vi è un "salto" pari alla fondamentale stessa. Possiamo parlare qui di *risoluzione di frequenza,* pari proprio al passo del pettine, cioè alla fondamentale: la nostra analisi "mescolerà" le frequenze che si trovino ad essere più vicine di questo passo, non permettendoci di risolverle, di distinguerle.

Più lungo l'intervallo di analisi, più alta la risoluzione di frequenza, più "fine" l'analisi. C'è qui un principio di conservazione, o se si vuole *un'invariante*: il prodotto tra risoluzione in frequenza e lunghezza dell'intervallo temporale di analisi è una costante (*un'invariante*) universale. Questo principio è detto anche *principio di indeterminazione,* perché è proprio formalmente nient'altro che l'espressione matematica, formale, del *principio di indeterminazione di Heisemberg* della meccanica quantistica.

Questa invarianza può essere vista come un'invarianza di un'area: l'area del rettangolo $\Delta T \cdot \Delta \nu$ è costante, e pari a $1/2$. Se allunghiamo una dimensione, dobbiamo restringere l'altra.

Dobbiamo notare una cosa importantissima, sia dal punto di vista concettuale, sia pratico (anche da un punto di vista musicale): il tempo relativo all'intervallo di analisi è un *tempo che si deve attendere* per avere il risultato dell'analisi stessa, se la stiamo effettuando in tempo reale, perché esso è il tempo necessario al fluire dei dati, ad ottenere l'ultimo dato necessario rispetto al primo.

Un analizzatore di spettro in tempo reale che operasse su di una base temporale *T* funzionerebbe così: appena avviato, non sarebbe in grado di produrre nessun dato in uscita. Trascorso il tempo *T*, sarebbe in grado di produrre lo spettro relativo all'intervallo appena trascorso. Da quel momento in poi, supponendo che l'operazione di produzione dello spettro sia istantanea, sarebbe in grado di produrre in tempo reale lo spettro del segnale, istante per istante riferentesi all'intervallo di tempo precedente, con un ritardo medio rispetto ai dati di $T/2$. Questo ritardo medio, e la latenza iniziale sono *ineliminabili* per ragioni di principio. Essi sono direttamente legati alla natura del tempo: ad ogni istante, noi possiamo conoscere solo il passato, quindi possiamo ottenere risultati che facciano uso dei valori passati[12].

Ecco dunque che il nostro principio di indeterminazione prende un significato particolarmente pregnante: per avere un'analisi dettagliata, bisogna attendere un tempo lungo. Analisi brevi producono invece informazioni grossolane. Ad esempio, per avere un'analisi con un dettaglio di 1 Hz, si dovrà aspettare ben un secondo: il nostro processo avrà dunque una latenza *intrinseca* (di principio) di un secondo, e un ritardo medio di $1/2$ secondo. Si tratta di una latenza *ineliminabile,* di natura concettuale, non matematica né tecnica: la si avrebbe anche se per eseguire i calcoli fosse necessario un tempo zero. Anzi, la latenza effettiva del processo di calcolo sarà la latenza intrinseca *più* il tempo necessario a completare i calcoli.

Nel discreto, come si vedrà nel capitolo apposito, di parla di DSTFT (*Discrete Short Time Fourier Transform*). A parte le specificità dovute alla discretizzazione del tempo, il discorso è del tutto analogo, salvo il fatto che nei sistemi numerici la trasformata non può essere un processo istantaneo, ma richiede un *tempo di calcolo,* dovuto al numero di operazioni (o di passi) necessario, e al fatto che tutto il sistema è discreto nel tempo, e la CPU è in grado di eseguire un determinato numero finito di operazioni al secondo.

Qualunque processo in tempo reale dell'elaborazione del suono nel dominio della frequenza deve quindi fare i conti con quest'invariante, con il "principio di indeterminazione" tempo-frequenza.

# 7 Sintesi additiva

Prendiamo ora in considerazione la sintesi additiva generica, con la quale si esegue la combinazione lineare di diversi oscillatori sinusoidali di frequenza arbi-

12 Questa osservazione è meno banale di quanto possa apparire a prima vista. E' infatti concepibile un processo "predittivo", nel quale si *stimino* i valori futuri a partire dalla conoscenza di quelli passati e da certe ipotesi ragionevoli sulla natura del sistema in esame. E' quello che fanno i filtri di Kalman, e a quanto pare, tutti i processi del vivente, i quali sembrano implicare in modo generale una "proiezione" nel futuro, una previsione (adattabile e correggibile), una "anticipazione" di quanto "sta per accadere".

traria. Nella pratica sia l'ampiezza, sia la frequenza degli oscillatori variano nel tempo (si tratta cioè di sistemi *tempo-varianti* (*time-variant*). Anche se si riferiscono a condizioni di invarianza nel tempo, le considerazioni qui svolte si adattano facilmente a casi più generali.

Non avendo in questo contesto più generale nessun vincolo sulle frequenze di sintesi, possiamo sceglierle in rapporti non interi, ad esempio irrazionali come $\sqrt{2}$, o addirittura trascendenti (come ad esempio $\pi$).

$$output(t)=\sum_{n=0}^{N} a_n \cdot \sin(\omega_n \cdot t+\varphi_n)$$

Limitiamoci all'inizio al caso di due sole frequenze, assumendo per semplicità in entrambi i casi la fase a zero:

$$output(t)=a_1 \cdot \sin(\omega_1 \cdot t)+a_2 \cdot \sin(\omega_2 \cdot t)$$

Poniamoci la seguente domanda: se le due frequenze sono tra loro in un rapporto qualunque, il segnale risultante è periodico? E in tal caso, qual è il suo periodo?

E' anzitutto evidente che in questa considerazione le due ampiezze $a_1$ e $a_2$ sono irrilevanti: le fasi delle due sinusoidi sono infatti a zero all'istante $t=0$, e la loro somma si ripeterà non appena le due fasi torneranno a definire contemporaneamente l'angolo zero (ovvero: saranno entrambe contemporaneamente un multiplo intero di $2\pi$), e questo indipendentemente dalle loro ampiezze. Chiamando *T* questo istante:

$\omega_1 \cdot T=n \cdot 2\pi$ e contemporaneamente $\omega_2 \cdot T=m \cdot 2\pi$

ovvero: $\nu_1 \cdot T=n$ e $\nu_2 \cdot T=m$ 12

con *n* e *m* interi. Dunque deve essere:

$$\frac{\nu_1}{\nu_2}=\frac{n}{m} \qquad 13$$

Ricordando che il rapporto tra due numeri interi è un numero detto *razionale*, abbiamo in altre parole:

***(III) Il segnale risultante dalla combinazione lineare di due sinusoidi è periodico se e solo se le loro frequenze sono in rapporto razionale.***

Questa considerazione ci permette di identificare i due termini: con "armonico" si intende un suono prodotto dalla combinazione lineare (*miscelazione)* di frequenze in rapporto razionale tra di loro.

E' facile generalizzare a più di due componenti:

$$\nu_1 \cdot T=n_1 \quad \nu_2 \cdot T=n_2 \ \ldots \ \nu_m \cdot T=n_m$$

dove gli $n_i$ sono tutti interi.

Supponiamo tanto per fissare le idee di avere ordinato le frequenze in ordine crescente, e che dunque $\nu_1$ sia la minore di queste.

Perché il segnale risulti periodico deve quindi essere contemporaneamente:

$$\frac{\nu_2}{\nu_1}=\frac{n_2}{n_1}\ \frac{\nu_3}{\nu_1}=\frac{n_3}{n_1}\ \frac{\nu_3}{\nu_1}=\frac{n_3}{n_1}\ \ldots\ \frac{\nu_j}{\nu_1}=\frac{n_j}{n_1}$$

***(IV) Il segnale risultante dalla combinazione lineare di più sinusoidi è periodico se e solo se le loro frequenze sono tutte in rapporto razionale armonico, appartengono cioè ad una sequenza armonica di una frequenza fondamentale.***

Per ottenere la frequenza fondamentale dalla sequenza ordinata è necessario esprimere i rapporti razionali con la minore di loro riducendoli ai minimi termini. Il più piccolo dei denominatori $n_{min}$ così ottenuti fornirà la fondamentale $\nu_f$ della sequenza:

$$\nu_f=\frac{\nu_1}{n_{min}}$$

Gli indici di armonica si otterranno a questo punto moltiplicando tutti i numeratori dei rapporti con la frequenza minore per $n_{min}$:

$$m_i=n_i\cdot n_{min}$$

Ottenendo una nuova sequenza, la sequenza armonica completa:

$$\nu_i=\nu_f\cdot i \quad \text{con} \quad i=0,1,\ldots \qquad 14$$

Se $n_{min}=1$ allora $\nu_f=\nu_1$ ovvero la fondamentale era presente nella sequenza originale (in tal caso, essa è $\nu-1$, la minore tra loro).

Dato che l'operazione indicata è sempre possibile su qualunque sequenza di razionali, ne discende che la combinazione lineare di frequenze in rapporto razionale tra di loro *è sempre periodica*.

Vale la pena di fare notare ancora una volta come in questo ragionamento non entri la considerazione delle ampiezze, ma solo delle frequenze. Alcune di queste frequenze armoniche, ivi inclusa come mostrato più sopra la fondamentale $\nu_f$, possono dunque avere ampiezza zero, cioè *non comparire nella sequenza*. Questo *non altera la periodicità*, che resta definita dalla $\nu_f$ anche qualora questa risulti assente. Come mostrato sopra, in tal caso $\nu_f$ non coinciderebbe con la frequenza più bassa presente nella sequenza, ma sarebbe la più bassa della sequenza armonica completa ottenuta dalla ricostruzione indicata in 14[13]. Si tratterebbe del fenomeno percettivo detto *fondamentale mancante (missing fundamental)*. E' noto che l'intonazione di un segnale sonoro è data dalla frequenza della fondamentale anche quando questa sia mancante.

In altre parole:

***(V) I segnali ottenibili da serie di Fourier sono tutti e soli i segnali periodici.***

Se si confronta quest'ultimo enunciato con I (pag. 5) si percepisce l'impor-

13 La più bassa ad eccezione della frequenza zero, *che fa sempre parte di ogni sequenza armonica*, com'è facile convincersi, eventualmente con ampiezza nulla.

tanza e il ruolo dell'analisi di Fourier, e il suo strettissimo legame con la periodicità del segnale (e quindi con i suoni intonati)[14].

Proviamo ora a risolvere il problema di determinare la periodicità di un segnale composto dalla combinazione lineare di due sinusoidi armoniche, date le loro frequenze. Una volta conosciuti *n* e *m*, una qualunque delle 12 ci fornisce *T*. Il problema di determinare la periodicità del segnale ottenuto dalla combinazione lineare di due sinusoidi, date le due frequenze $\nu_1$ e $\nu_2$ è dunque ricondotto a quello di esprimere il loro rapporto come *numero razionale fratto ridotto ai minimi termini*[15], il che significa – ricordiamolo dalle medie inferiori – esprimerlo in un numero razionale qualunque e poi dividere numeratore e denominatore per il loro *massimo comune divisore* (MCD).

A titolo di esempio si considerino le seguenti due frequenze:

$$\nu_a = 4076.48\,Hz \qquad \nu_b = 5606.16\,Hz$$

Abbiamo: $\dfrac{\nu_b}{\nu_a} = 1.375 = \dfrac{11}{8}$

Dunque: $\nu_a = 8 \times 509{,}56\,Hz$ e $\nu_b = 11 \times 509{,}56\,Hz$ .

E' del tutto evidente che le due frequenze sono rispettivamente la 8ª e la 11ª armonica di $\nu_1 = 509{,}56\,Hz$ , la quale fornisce dunque la *periodicità* (in questo caso, anche l'intonazione) del suono risultante.

## *7.1 Periodicità nei sistemi numerici.*

I sistemi numerici (detti anche digitali) fisicamente realizzabili sono non solo *discreti* (costituiti da un numero contabile di parti e dotati di un numero contabile di stati), ma anche *finiti*. Questo indica che non solo il tempo è discreto, ma che anche le ampiezze campionate sono discrete e finite: in altre parole, abbiamo una rappresentazione delle grandezze *intera e finita.*

Nella memoria dei calcolatori dunque i numeri sono sempre interi. I loro

14 Vale la pena di notare che in tutto questo ragionamento non solo non interviene la considerazione delle ampiezze, ma neanche della *forma* delle funzioni usate come base per comporre il segnale, ma solo della loro periodicità. Le stesse considerazioni resterebbero valide assumendo come segnali elementari *funzioni periodiche qualsivoglia,* purché con periodicità in sequenza armonica. Queste considerazioni non determinano quindi nessun nesso tra intonazione, intesa come periodicità, e sinusoidi, e quindi non determinano nessun ruolo privilegiato delle sinusoidi tra tutte le funzioni periodiche immaginabili. La peculiarità delle sinusoidi sta in loro proprietà matematiche che le rendono a loro volta soluzioni di determinate equazioni in grado di descrivere alcuni sistemi fisici fondamentali, *gli oscillatori.* Questi sono il meccanismo fisico sottostante agli oggetti utilizzati come strumenti musicali, ivi incluso il nostro apparato fonatorio. *Il ruolo privilegiato delle funzioni sinusoidali è quindi da ricercarsi in ultima analisi nella fisica, e nella fisiologia umana, non nella matematica.* E' del tutto concepibile un'analisi-sintesi armonica fatta a partire da funzioni periodiche diverse dalle sinusoidi, ad esempio delle funzioni triangolari, o quadre. Analisi-sintesi del genere avrebbero però scarsi punti di contatto con il nostro sistema uditivo.

15 Questo perché, come si è detto, se un segnale è periodico con periodo *T,* allora lo è anche con periodo 2*T*, 3*T*, ..., n*T*. Questi ultimi sono cioè tutti *possibili periodi* per il segnale, ma noi siamo interessati al più piccolo di loro, cioè *T*. E dunque dobbiamo trovare i *minimi termini* del rapporto razionale.

rapporti sono quindi sempre *razionali*. Anche quando si tratti di esprimere un numero irrazionale (come $\sqrt{2}$) o trascendente (come $\pi$), è giocoforza adottarne una *approssimazione razionale*. Come conseguenza, la combinazione lineare di funzioni periodiche qualunque di frequenze esprimibili in un sistema numerico *è sempre periodica*[16]. Da questa affermazione sembrerebbe scaturire che con i sistemi numerici è possibile ottenere solo suoni intonati. Si tratta però di una affermazione erronea, senza che vi sia nessuna contraddizione con la forzata periodicità di cui si sta parlando. Infatti un suono periodico è intonato (categoria quest'ultima percettiva, non matematica, come già detto) se il suo periodo ricade nella banda audio: 20÷20.000 Hz. Segnali con periodicità più breve sono semplicemente non udibili (sono *ultrasuoni,* ancorché intonati, circostanza della quale potrebbero godere diversi animali quali cani, gatti e uccelli, ma non gli esseri umani). Segnali con periodicità più lunga dei 20 Hz vengono invece avvertiti come dotati di una "prosodia periodica", ma non necessariamente intonati. Se la periodicità è inoltre molto lunga (dell'ordine dei minuti, delle ore o addirittura degli anni) può anche non essere più avvertibile come tale, perché al di fuori dalle nostre capacità temporali di memorizzazione.

Ad esempio, si considerino le due frequenze:

$$\nu_a = 63{,}41958396752917\,Hz \qquad \nu_b = 87{,}20192795535262\,Hz$$

Abbiamo: $\dfrac{\nu_b}{\nu_a} = 1{,}375000000125 = \dfrac{11.000.000.001}{8.000.000.000}$

Dunque:

$$\nu_b = 11.000.000.001 \cdot 7{,}972\ldots \times 10^{-9}\,Hz \quad \text{e} \quad \nu_a = 8.000.000.000 \cdot 7{,}972\ldots \times 10^{-9}\,Hz$$

In altre parole, sulle medesime tracce di quanto fatto già nel cap. 7, riconosciamo che le due frequenze altro non sono rispettivamente che la 11 miliardesima e uno, e la 8 miliardesima armonica della frequenza:

$$\nu_1 = 7{,}927447995941146 \times 10^{-9}\,Hz$$

Questa frequenza definisce la periodicità $T = 1/\nu_1$ del segnale risultante. Si tratta di una frequenza molto bassa, alla quale corrisponde dunque un periodo lungo. Ma quanto lungo? Il conto è presto fatto: si tratta esattamente di quattro anni.

Ci si potrebbe domandare a questo punto se nella pratica, con i calcolatori reali, risulti possibile ottenere delle periodicità così lunghe, ovvero se queste siano compatibili con la precisione numerica disponibile e con i corrispondenti arrotondamenti. La risposta è sì, e anche con una certa comodità. Una frequenza del genere richiede 16 cifre decimali, mentre il calcolatore con il quale il presente testo è stato scritto ne fornisce 17.

Un'altra domanda che ci si potrebbe porre è se la circostanza che oggi i calcolatori elettronici utilizzano la rappresentazione (e memorizzazione) dei nume-

16 Il risultato è molto più generale: qualunque algoritmo di una macchina finita produce sequenze di numeri (segnali) solo periodiche. Ma anche per queste valgono le considerazioni svolte poco oltre nel testo.

ri in virgola flottante (*floating point)* modifichi in qualche modo questi risultati. La risposta è no. La notazione in virgola flottante è solo un modo (migliore) di gestire la rappresentazione intera, in modo tale da distribuire meglio l'errore di troncamento, di effettuare cioè uno scalamento automatico della rappresentazione in modo da rendere il “quanto” (la granularità) proporzionale al valore assoluto della grandezza stessa rappresentata. La notazione in virgola flottante continua comunque ad approssimare irrazionali e trascendenti con numeri razionali, un destino questo direttamente legato alla finitezza dei registri di memoria, al quale non c'è modo di sfuggire. Resta sempre il fatto che il numero trascendente $\pi$ è approssimato con un razionale:

$$\pi = 3{,}141592653589793 = \frac{3.141.592.653.589.793}{1.000.000.000.000.000}$$

Un segnale ottenuto dalla somma di una sinusoide a 20 Hz e di una a $\pi \cdot 20\,Hz$ in una rappresentazione come quella qui indicata avrebbe una periodicità di 6.341.958.397 anni, un periodo paragonabile con il restante tempo del ciclo dell'idrogeno del sole. In breve, con la restante vita del nostro sole così come lo conosciamo.

### *7.2 Armonicità e anarmonicità.*

Le cose vanno però un po' diversamente se si vogliono considerare non solo i segnali periodici, ma anche quelli quasi-periodici, in considerazione del fatto che alcuni di questi sono percepiti come intonati. In questo caso le ampiezze possono rientrare in gioco. Se ad esempio costruiamo una combinazione lineare di segnali sinusoidali nella quale un sottoinsieme di frequenze tra loro armoniche abbia una ampiezza considerevole, e siano presenti alcune frequenze inarmoniche con ampiezza piccola, otterremo un segnale quasi-periodico, con un periodo percepito *T* definito dal sottoinsieme armonico. La periodicità di questo segnale risultante sarà solo lievemente perturbata dalla presenza delle (piccole) componenti non armoniche. L'effetto di queste ultime è assimilabile ad un *jitter* di frequenza (ma in questo caso non casuale), ottenendo quindi un suono intonato con una caratteristica timbrica dovuta alle componenti inarmoniche.

E' il caso delle corde rigide, come sono ad esempio quelle basse del pianoforte. La loro rigidità (elasticità flessoria) conferisce al suono un caratteristico timbro “metallico”, dovuto a componenti anarmoniche. L'anarmonicità cresce con l'indice di parziale[17] con un andamento del tipo:

$$\nu_n = \nu_1 \cdot \sqrt{n^2 + \alpha n^4}$$

dove $\alpha$ è proporzionale alla rigidità, ma è comunque, in condizioni reali, un numero molto piccolo. Come conseguenza, le prime parziali sono molto prossime all'armonicità (nella formula, se si pone $\alpha = 0$ si ottiene un pettine armonico). Se a questo si aggiunge che le parziali alte, che sono più anarmoniche, hanno ampiezze piccole anche come conseguenza del fatto che si smorzano più rapidamente, è facile intendere come un suono del genere risulti fondamentalmente

17 E' corretto usare in questo caso il termine “parziale”, dato che i rapporti tra le frequenze non sono più rigorosamente armonici.

intonato, e solo “colorato” dalla presenza di parziali alte inarmoniche.

L'anarmonicità qui descritta è “fisica”, tipica appunto delle corde rigide (metalliche). Nella musica elettronica si è fatto uso di un tipo di anarmonicità più accentuata e equamente distribuita (anziché concentrata nelle alte frequenze) e di natura prettamente matematica. Non esistono sistemi fisici meccanici che riproducano un comportamento del genere. L'espressione di questo tipo di anarmonicità è la seguente:

$$\nu_n = \nu_1 \cdot n^B$$

dove *B* è detto *indice di anarmonicità* o *indice di Nichols*.

Per $B=1$ abbiamo una sequenza perfettamente armonica. Per $B>1$ abbiamo una sequenza “superarmonica”, nella quale le distanze frequenziali sono superiori a quelle di una sequenza armonica (il pettine si allarga progressivamente all'aumentare della frequenza). Per $B<1$ abbiamo una sequenza “sottoarmonica”, nella quale le distanze tra frequenze sono inferiori a quelle armoniche (il pettine di infittisce all'aumentare della frequenza)[18].

Che tipo di suoni si ottengono da sequenze siffatte? In particolare, si ottengono suoni periodici e intonati, e sotto quali condizioni?

E' ovviamente possibile escogitare qualunque tipo di andamento, ad esempio polinomiale, esponenziale, ecc., monotono o non monotono. Se non vi sono componenti prevalenti in termini di ampiezza con rapporti armonici o quasi-armonici, ci si deve attendere un suono non intonato, e assai presumibilmente dotato di una marcata “prosodia” priva di ripetizioni (dato che il segnale risultante non è periodico) dovuta ad un complesso gioco di “battimenti”.

### *7.3 Suoni composti*

Un suono può essere composto dalla somma di due o più suoni armonici. E' quello che musicalmente è detto un *accordo*[19]. Visto come un suono unico, esso in generale non sarà periodico[20]. E' però possibile identificare le diverse fondamentali presenti operando su tutte le coppie possibili di frequenze presenti. Un'operazione del genere è ostacolata, partendo da suoni concreti, dalla imprecisione della determinazione delle frequenze a causa sia degli errori di misura, sia del fatto che le frequenze di partenza non sono in genere matematicamente esatte. Tra le cause degli errori di misura della frequenza dobbiamo annoverare l'onnipresenza, in qualche misura, di *rumore di fondo*, dalle origini più svariate (ambientali, elettroniche, elettroacustiche), che inevitabilmente si somma (si *miscela)* al segnale originale.

18 Nei grafici spettrali uno spettro armonico appare equispaziato solo quando si scelga una scala lineare per le frequenze, cosa che non è la scelta usuale. Il classico grafico bilogaritmico presenta i pettini armonici non equispaziati, ma infittentisi verso le alte frequenze.

19 Vale la pena di fare notare che non si fa qui questione se si tratti di un accordo “consonante” o dissonante”. Per *accordo* si intende qui la presenza contemporanea di più note distinte qualsivoglia.

20 Qui si cela dietro le scene la questione dei “temperamenti”. Nel temperamento equabile nessun accordo fornisce suoni periodici, ad eccezione dell'ottava e suoi multipli, dato che nessun intervallo è “giusto” - cioè razionale - salvo l'ottava.

### 7.4 Anarmonicità e anarmonicità nei sistemi numerici

Le considerazioni fin qui svolte ci permettono di fornire una definizione sensata di “suono anarmonico” (o inarmonico) nei sistemi numerici (che, ricordiamolo, producono solo suoni periodici e quindi, a rigore, solo suoni matematicamente armonici).

***(VI)*** ***Definiamo come “anarmonico” in un sistema numerico un suono (armonico) che abbia la fondamentale al di sotto della banda audio, e che non sia riconducibile alla somma di più suoni armonici in senso numerico (aventi cioè la fondamentale in banda audio). Se si trattasse invece di un suono riconducibile alla somma di più suoni armonici, si tratterebbe di un accordo.***

## 8 La trasformata di Fourier numerica: DFT e FFT.

Nei sistemi a tempo discreto (campionati), come noto, le frequenze rappresentabili sono limitate superiormente alla *frequenza di Nyquist* (pari alla metà della frequenza di campionamento). La trasformata di Fourier calcolata in un sistema discreto fornirà quindi una serie armonica limitata superiormente a Nyquist, quindi una *serie finita* di frequenze. Nel continuo, invece, abbiamo una *serie infinita* di frequenze discrete.

Il calcolo della DFT è in realtà concettualmente semplice. Stabilito il numero *N* di campioni $y_0, y_1, \ldots, y_{N-1}$ presi in esame ai tempi $t_0, t_1, \ldots, t_{N-1}$, è possibile impostare un sistema di *N* equazioni indipendenti, e determinare quindi *N* coefficienti. Poiché ogni frequenza richiede due coefficienti, si possono determinare *N/2* frequenze.

Il sistema di *N* equazioni è semplicemente:

$$\sum_{n=0}^{\frac{N}{2}-1} a_n \cdot \cos(\omega_n \cdot t_j) + b_n \cdot \sin(\omega_n \cdot t_j) = y_j \quad \text{con} \quad j=0,\ldots,N-1 \qquad 15$$

Si noti che si può (e conviene) assumere zero come origine dei tempi, per cui:

$$t_j = j \cdot \Delta T$$

dove $\Delta T$ è il periodo di campionamento.

Dato che il numero di frequenze determinabili è *N/2,* e dato il limite di Nyquist, anche le frequenze sono determinate:

$$\omega_n = \frac{n \cdot \pi}{N \, \Delta T}$$

Queste frequenze definiscono un pettine equispaziato (armonico) di passo $\Delta \nu = \frac{1}{\Delta T}$ (e quindi di fondamentale $\Delta \nu$ ). Anche qui abbiamo una *risoluzione in frequenza* dell'analisi: essa è tanto maggiore (tanto più dettagliata l'analisi) quanto più lungo l'intervallo di tempo di analisi.

Torniamo ora al calcolo numerico della trasformata. Nel caso discreto la finitezza del numero di armoniche trasforma il calcolo degli integrali 10 e 11 (pag. 10) nella risoluzione di un sistema di equazioni lineari.

I coseni e seni nella 15 costituiscono una matrice di $N \cdot N/2 = N^2/2$ coefficienti fissi, che dato *N* possono essere calcolati una volta per tutte:

$$C_{n,j} = \cos(\frac{j \cdot n \cdot \pi}{N}) \quad S_{n,j} = \sin(\frac{j \cdot n \cdot \pi}{N})$$

E quindi la 15 può essere riscritta:

$$\sum_{n=0}^{\frac{N}{2}-1} a_n \cdot C_{n,j} + b_n \cdot S_{n,j} = y_j \qquad 16$$

$$(a_0 \cdot C_{0,0} + b_0 \cdot S_{0,0}) + (a_1 \cdot C_{1,0} + b_1 \cdot S_{1,0}) + \ldots + (a_N \cdot C_{N/2,0} + b_N \cdot S_{N/2,0}) = y_0$$
$$(a_0 \cdot C_{0,1} + b_0 \cdot S_{0,1}) + (a_1 \cdot C_{1,1} + b_1 \cdot S_{1,1}) + \ldots + (a_N \cdot C_{N/2,1} + b_N \cdot S_{N/2,1}) = y_2$$
$$\ldots\ldots\ldots\ldots$$
$$(a_0 \cdot C_{0,N} + b_0 \cdot S_{0,N}) + (a_1 \cdot C_{1,N} + b_1 \cdot S_{1,N}) + \ldots + (a_N \cdot C_{N/2,0} + b_N \cdot S_{N/2,0}) = y_N$$

La 16 definisce un *sistema di equazioni lineari*, un tipico oggetto di interesse della *algebra lineare (linear algebra)*. La sua soluzione (dati gli $y$, e conosciute le costanti $C$ ed $S$*,* trovare gli $a$ e $b$*)* richiede la *inversione di una matrice,* un'operazione banale in linea di principio ma che comporta notevoli problemi di quantità di calcoli non appena questa sia di dimensioni non minuscole. Ed è questo il caso dei segnali audio, dato che è del tutto normale che $N$ sia dell'ordine delle centinaia o anche delle migliaia.

Ad esempio, si supponga di volere fare l'analisi di Fourier di un frammento di suono della durata di un secondo, il che significa adottare uno spettro armonico che ha come fondamentale 1 Hz (e dunque di passo 1 Hz). Se il frammento è fornito in standard CD, è costituito di 44.100 campioni, e dunque $N = 44.100$ .

Altro esempio: un tipico spettrogramma adatto alle caratteristiche del sistema uditivo umano è eseguito su intervalli temporali da 10 a 20 msec, il che implica, sempre nello standard CD, da 441 a 882 campioni ( $N = 441 \div 882$ *).*

Il motivo di un intervallo del genere sta nella circostanza che il sistema uditivo umano non distingue una prosodia (una sequenza temporale di eventi) che sia più rapida di questi tempi. Un suono di questa durata è percepito come un unico evento sonoro (un *atto* sonoro) senza struttura temporale interna. In termini dell'analisi di Fourier, questo significa che, su di una scala di tempo del genere[21], il nostro sistema uditivo *non è sensibile all'informazione di fase*, ovvero che *suoni con identici profili spettrali ma differenti profili di fase vengono percepiti come lo stesso suono*[22]. Fornire la sola informazione di ampiezza su questa

21 Non si sottolineerà mai abbastanza che l'affermazione che segue è valida solo sotto questa ipotesi, e solo per il *nostro* sistema uditivo.

22 Come tutte le affermazioni relative al vivente e alla percezione, anche questa deve essere presa in modo non categorico. In determinate condizioni possono essere percepibili differenze minori. Infatti le fasi giocano un ruolo nella determinazione del *fattore di cresta* del segnale risultante, ovvero del rapporto tra ampiezza di picco (massima) e ampiezza media, e talvolta

scala di tempo ha dunque senso, perché l'informazione di fase è percettivamente irrilevante. Va però precisato che questo ordine di grandezza è valido per gli esseri umani e alcuni animali, ma non per tutti gli animali. In genere, l'intervallo temporale ad di sotto del quale si perde la capacità di distinguere le sequenze temporali di eventi sonori dipende grosso modo dalle dimensioni dell'orecchio interno e quindi – sempre grosso modo – dalle dimensioni della testa. Più piccole le dimensioni, più breve l'intervallo. Gli uccelli, ad esempio, hanno una percezione più fine degli eventi temporali, e il loro limite si situa attorno al millisecondo. Le balene, viceversa, hanno intervalli tipici di centinaia di millisecondi. Gli studi effettuati sulla percezione uditiva degli uccelli ci dicono che noi non possiamo comprendere il loro cinguettio (che ha funzioni comunicative) perché il messaggio sta, più che negli aspetti melodici, nella prosodia (non diversamente da quanto accada nelle lingue non ideogrammatiche). Ma questa negli uccelli si sviluppa ad una scala di tempo molto inferiore alle nostre capacità di discriminazione. Noi dunque potremmo forse comprendere il cinguettio degli uccelli, se solo fossimo in grado di percepirlo come loro lo percepiscono, ma questo ci è negato perché mescoliamo eventi che per loro sono distinti e costituiscono una sequenza percepibile[23]. Per gli stessi motivi, le balene non possono percepire il parlato umano.[24]

*Quando si parla di DFT, si intende in genere il calcolo della trasformata discreta di Fourier eseguito in modo diretto*, invertendo la matrice caratteristica del sistema di equazioni. Questo calcolo comporta un numero di operazioni elementari (somme e moltiplicazioni) proporzionale a $N^2$ . E' facile intendere che per i tipici $N$ qui citati, il numero di operazioni elementari diventa enorme $(44.100^2 = 1.944.810.000)$ . Nel gergo del calcolo automatico si dice che la *complessità di calcolo* di questo algoritmo è pari a $O(N^2)$ .

Il compito di calcolare i coefficienti dello sviluppo di Fourier non può però essere affidato ad un meccanismo decisamente inadatto alle esigenze pratiche come la DFT. Già Gauss[25] si era accorto che era possibile sfruttare determinate proprietà di simmetria dei calcoli da eseguire per ridurne il numero. In particolare, la trasformata di Fourier di $N$ campioni può essere ridotta al calcolo di due trasformate (più semplici) di $N/2$ campioni, purché (ovviamente) $N$ sia pari. Queste a loro volta, se $N/2$ continua ad essere pari, possono essere ricondotte a 4 trasformate di $N/4$ campioni, e così via, giungendo fino a trasformate di due soli campioni. Perché questo sia possibile, $N$ deve essere indefinitamente

differenze nel fattore di cresta possono dare luogo ad una differente intensità percepita.

23 Potremmo però utilizzare la tecnica del *time-stretching* (dilatazione temporale), che altera il tempo (in questo caso, lo allunga) senza alterarne la composizione spettrale. Qualcuno ha chiamato questa tecnica "microscopio acustico", un microscopio che opera "ingrandendo" il tempo.

24 E' questo quel che intendeva il compositore spettralista Gérard Grisey quando parlava del "tempo degli uccelli" e del "tempo delle balene".

25 Carl Friedrich Gauss (1777-1855) è stata una figura di importanza capitale nella storia del pensiero. Fondamentali i suoi lavori sui numeri complessi e su diversi metodi di calcolo, che applicò alla disciplina alla quale teneva più di tutte: l'astronomia. A lui si devono, anche se mai pubblicate, le prime intuizioni sulle geometrie non euclidee e sulla possibilità che il nostro spazio sia in realtà curvo, anticipando - in questo – le idee successive di Riemann e anche quelle della Relatività generale di Albert Einstein.

divisibile per due: non basta dunque che sia pari, ma deve essere una *potenza di due:* $N = 2^M$.[26] Operando in questo modo (una sorta di *divide et impera*), la complessità si riduce a $O(N \cdot \log_2(N))$. Ricordando che il logaritmo in base 2 fornisce il numero di cifre binarie necessarie ad esprimere l'argomento, è facile convincersi della enorme differenza tra questo metodo e quello diretto DFT. Anzitutto, dato che *N* deve essere pari ad una potenza di due, $\log_2(N)$ è un numero intero. Eseguiamo il calcolo per *N=65.536* (che è notevolmente maggiore di 44.100): $65.536 \cdot 16 = 1.048.576$, da confrontare con i quasi due miliardi del paragrafo precedente.

Nel 1965 J.W. Cooley (della IBM) e John Tukey (della *Princeton University*) "riscoprirono" le idee di Gauss applicandole al calcolo automatico: vide così la luce l'algoritmo FFT (*Fast Fourier Transform*), che avrà una capitale importanza in tutta l'elaborazione del segnale.

Anche per questo algoritmo esistono sia "sistemi prefabbricati" che librerie software[27]. Successivamente si sono trovati algoritmi efficienti sotto il vincolo che *N* sia un numero primo "piccolo", o sia esprimibile come prodotto di potenze di numeri primi piccoli. Questo approccio è contenuto nella libreria *Open Source* FFTW, altamente ottimizzata, che fornisce diversi tipi di trasformate in un numero arbitrario di dimensioni.[28]

### *8.1 Antitrasformata.*

L'operazione contraria alla trasformata, ottenere cioè il segnale nel tempo a partire dai coefficienti della trasformata di Fourier, prende il nome di *antitrasformata,* e si indica spesso con $FFT^{-1}$. L'esponente *-1* è sovente usato in matematica per indicare "operazione inversa" in analogia alla moltiplicazione:

$$a^{-1} = \frac{1}{a} \quad \text{dalla quale:} \quad a \cdot a^{-1} = 1 \qquad 17$$

Anche nel nostro caso $FFT^{-1}(FFT)$ o anche $FFT(FFT^{-1}) = 1$ o, in altre parole, l'antitrasformata di una trasformata è un'operazione unitaria: è pari infatti al segnale di partenza (così come la trasformata di un'antitrasformata è uguale allo spettro di partenza).

Ricordiamo che la trasformata conduce dal dominio del tempo a quello della frequenza; l'antitrasformata conduce dal dominio della frequenza a quello del tempo.

Per calcolare l'antitrasformata si può ricorrere alla formula esplicita 15 o 16.

26 Ad esempio: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1.024, 2.048, 4.096, 8.182, 16.384, 32.768, 65.536.

27 Una per tutte: la libreria GNU *Open Source* GSL (*Gnu Scientific Library*), disponibile sotto forma di sorgente in linguaggio C, che contiene anche le funzioni dell'algebra lineare. La libreria è altamente collaudata e ottimizzata. (http://www.gnu.org/software/gsl/).

28 "FFTW is best at handling sizes of the form $2^a\ 3^b\ 5^c\ 7^d\ 11^e\ 13^f$, where e+f is either 0 or 1, and the other exponents are arbitrary. Other sizes are computed by means of a slow, general-purpose algorithm (which nevertheless retains *O*(*n* log *n*) performance even for prime sizes)" - http://www.fftw.org/

La complessità è anche in questo caso $O(N^2)$ [29] dove *N* è il numero di frequenze (armoniche). Ricordiamo che otterremo $2N$ campioni nel tempo. Il calcolo dei seni e dei coseni può essere effettuato una volta per tutte fuori linea (i coefficienti non cambiano, se non cambia il numero di punti), e il loro costo di calcolo può dunque essere trascurato.

Anche per l'antitrasformata si possono adottare scorciatoie simili a quelle della trasformata, e ottenere un algoritmo veloce del tutto simmetrico alla FFT, con complessità $O(N \cdot \log_2(N))$. Ovviamente, i vincoli sono gli stessi della FFT: *N* deve essere una potenza di due, oppure obbedire a schemi implicanti numeri primi piccoli.

## *8.2* FFT di un numero di campioni arbitrario: lo *zero-padding*.

E' legittimo chiedersi come fare quando l'intervallo temporale sul quale si desidera effettuare l'analisi di Fourier non sia costituito da un numero di campioni pari ad una potenza di due.

Esistono altri algoritmi "rapidi", come già detto, che operano sotto ipotesi diverse, ma non esiste un "algoritmo rapido" per *N* qualunque.

Lo *zero-padding* fornisce una soluzione: accettando un certo allungamento dei calcoli, si può procedere aggiungendo in coda, prima dell'analisi, un numero di campioni nulli in numero tale da raggiungere la più prossima potenza di due. Ovviamente ciò comporta il fatto che l'analisi che si ottiene non è quella del segnale originale, ma quella del segnale "allungato". La conseguenza inevitabile, e più importante, è che la struttura delle armoniche (cioè la fondamentale) sarà basata sulla nuova lunghezza. Sappiamo tuttavia da quanto detto nel cap. 6 a pag. 17 che questo non cessa di dare informazioni sulla "struttura sottostante", perché se il segnale "conteneva" delle sinusoidi "immerse", lo spettro si "addenserà" attorno alle loro frequenze. Abbiamo anche detto sempre nel cap. 6 che questa informazione tanto meglio viene evidenziata, quanto più lungo è il periodo di analisi, e noi facendo lo *zero-padding* stiamo effettivamente allungando la base temporale rispetto a quella iniziale. Stiamo, in altre parole, *migliorando formalmente* la risoluzione di frequenza.

Inoltre, inserire campioni nulli equivale a sommare un segnale nullo per tutto il periodo di analisi. Ma un segnale nullo ha componenti armoniche nulle, quindi l'operazione non perturba (salvo che per il fatto stesso di avere allungato la finestra d'analisi) lo spettro preesistente.

Si deve tuttavia precisare che il miglioramento della risoluzione in frequenza è *formale,* nel senso che avremo più punti della nostra analisi, ma non *sostanziale,* nel senso che questo non aumenta la nostra capacità di discriminare frequenze molto vicine. Infatti la durata del segnale effettivo è rimasta invariata, e pertanto la determinazione delle frequenze in esso contenute (la larghezza delle righe), e quindi la *precisione in frequenza* resta la stessa: semplicemente ogni riga allargata (dalla finitezza dell'intervallo temporale della sua durata) verrà

29 Si tratta infatti di eseguire *N/2* volte un somma di due prodotti per ottenere un punto sui 2*N* totali. Quindi $N/2 \cdot 2\mathrm{N}$.

descritta con più punti, ma la sua larghezza resterà la medesima.

## 9 Analisi-resintesi di Fourier

Da quanto abbiamo detto nel par. 8.1 possiamo concepire un'operazione consistente nell'eseguire la trasformata di Fourier di un segnale e poi, in catena, nella sua antitrasformata. L'operazione è concepibile (e perfettamente realizzabile) ma avrebbe poco senso: otterremmo infatti nient'altro che una esatta copia del segnale originale salvo, se l'operazione fosse compiuta in tempo reale, un ritardo tra i due pari almeno all'intervallo temporale di analisi – un risultato che sarebbe possibile ottenere con mezzi molto più immediati e semplici, come ad esempio una linea di ritardo.

Tuttavia il fatto che la trasformata di Fourier ci permette di guardare al segnale sotto un altro punto di vista (anche se del tutto equivalente, dal punto di vista matematico) può rendere facilmente effettuabili[30] delle *trasformazioni* del segnale nel dominio della frequenza prima dell'antitrasformazione. Notiamo infatti la dualità delle due rappresentazioni:

- ***(VII) Nel dominio del tempo le frequenze si presentano tutte "mescolate", inseparabili tra di loro: ad ogni campione nel tempo contribuiscono tutte le frequenze d'analisi. In compenso, il tempo è perfettamente separato in istanti successivi, ordinati.***
- ***(VIII) Nel dominio delle frequenze è il tempo (all'interno dell'intervallo di analisi) che si presenta "mescolato": ad ogni campione di frequenza contribuiscono tutti gli istanti di tempo. In compenso, le frequenze sono tutte separate tra di loro e ben ordinate.***

Nel dominio della frequenza possiamo ad esempio alterare il profilo spettrale del segnale, sopprimendo o attenuando, oppure esaltando o introducendo determinate frequenze o zone di frequenza. L'operazione può però risultare meno immediata di come possa sembrare a prima vista, proprio per quanto affermato in VIII: non è né facile né immediato capire, prevedere, cosa accada nel dominio del tempo modificando un profilo spettrale. Si tratta di tenere presenti due fenomeni, tra i quali il primo già citato:

- ***Le fasi delle armoniche governano in modo essenziale la prosodia del segnale, in modo piuttosto complesso.***
- ***Anche lasciando inalterate le fasi[31], la modificazione dei profili spettrali deve tenere conto del "fenomeno di Gibbs", di cui si parlerà tra breve.***

Vi sono però casi, circostanze (e finalità), nelle quali – operando con le cautele e le procedure richieste dal caso, e caso per caso – la modifica del segnale nel dominio della frequenza è particolarmente adeguata agli scopi (o addirittura

30 Vi sono operazioni, come ad esempio la sottrazione spettrale con soglia, quale quella effettuata nell'eliminazione del rumore, che sono fattibili solo nel dominio della frequenza.

31 Se però si vogliono introdurre frequenze assenti (presenti con ampiezza zero) nel segnale originale si è costretti a scegliere una fase, la quale avrà conseguenze sulla prosodia. Non ci si illuda che la "fase zero" rappresenti una scelta "neutra": zero è un numero come ogni altro.

l'unica possibile), o rappresenta comunque una valida possibilità. Citiamo tra i tanti due esempi: l'alterazione dell'intonazione o la contrazione o dilatazione del tempo (*pitch shift* e *time stretching*), che possono essere effettuate proficuamente sia nel dominio del tempo sia della frequenza[32], e la soppressione del rumore di fondo (mediante sottrazione spettrale).

Se si utilizza per l'analisi-alterazione-resintesi una base temporale breve (i 10÷20 msec di cui si parla nel cap. 8 a pag. 29) si cumulano un paio di vantaggi:

- Il tempo di latenza e il ritardo medio sono brevi.
- L'informazione di fase non è rilevante.

Come contropartita, la risoluzione di frequenza è piuttosto grossolana: 100÷200 Hz. Se questa risoluzione è sufficiente, si può pensare ad esempio di alterare il profilo spettrale per questa via, risintetizzando il segnale alterato nel dominio della frequenza mediante antitrasformazione. Se si trascura però la fase (ad esempio forzandola a zero) ci si trova di fronte alla circostanza di produrre un segnale periodico (quindi intonato) alla frequenza di 100÷200 Hz, perché tra una frame e la successiva si perderebbe la *continuità di fase,* che invece ha un rilievo percettivo primario, perché è il meccanismo che permette la creazione di frequenze che non appartengono al pettine delle armoniche di analisi (quindi, frequenze ad esempio inferiori alla frequenza di analisi) permettendo attraverso i concatenarsi delle frame il costituirsi di periodicità (e non periodicità) qualsivoglia.

L'alterazione pura e semplice dei profili spettrali si fa però meglio (e con latenze minime) nel dominio nel tempo, con i filtri numerici, che saranno oggetto di trattazione specifica. I filtri IIR (*Infinite Impulse Response*) sono particolarmente adeguati alle applicazioni musicali.

Volendo procedere nel dominio della frequenza, un modo per preservare la continuità di fase senza occuparsene direttamente è quello di operare con la usata (ed abusata) tecnica della *somma e sovrapposizione* (*overlap and add*).

### *9.1 La tecnica "overlap and add" (OLA): introduzione*

Un modo per antitrasformare (risintetizzare) un segnale su intervalli temporali (*frame)* arbitrari mantenendo continuità di fase è la tecnica "*overlap and add*", espressione che dà luogo all'acronimo *OLA*.

In pratica, si opera come in figura.

32 Sembra che gli algoritmi nel dominio della frequenza portino a risultati pratici superiori.

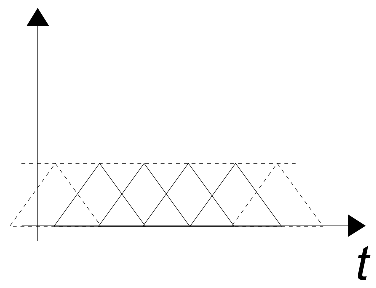


Fig. 11 - Tecnica *overlap and add:* la somma di triangoli sovrapposti è una costante

Si opera come nella STFT, ma per frame sovrapposte (tipicamente: al 50%). Si altera il segnale nel modo voluto, e si antitrasforma. Ogni *frame* viene però miscelata con la seconda metà della precedente e la prima metà della seguente, tutte e tre moltiplicate per un inviluppo triangolare. Operando in questo modo non si altera l'ampiezza del segnale (la somma dei tratti triangolari crescente e decrescente è costante) e per forza di cose la fase del segnale risultante è continua. Si tratta in pratica di un doppio *cross fade* lineare.

In questo modo il ritardo medio raddoppia, e diventa pari alla lunghezza della frame (in pratica, bisogna accumulare due frame prima di tirare fuori una nuova frame).

La tecnica *overlap and add* permette di giustapporre frame sonore senza produrre discontinuità sui bordi, lasciandone inalterato il contenuto di frequenze (almeno in senso relativo). Per questo è utilizzata in moltissime applicazioni, (ivi inclusa la sintesi granulare), per alterare la durata di segnali senza alterarne l'intonazione (o viceversa), come accade nella sintesi a *wave-tables* o a campionamento (la tecnica di sintesi corrente negli attuali *expander*) ed anche nella sintesi della voce.

L'inviluppo triangolare è un caso particolare di *finestra* nel dominio del tempo. Possiamo esprimere matematicamente l'operazione di applicazione dell'inviluppo come segue:

$$output(t)=w(t)\cdot input(t)$$

dove $w(t)$ indica appunto una finestra (*window).* Se la finestra è lunga *T,* l'operazione *overlap and add* può essere scritta:

$$output(t)=w(t)\cdot input_1(t)+w(t-T/2)\cdot input_2(t)+...+w(t-nT/2)\cdot input_n(t)$$

dove le *input* sono le varie frame, che si suppone siano nulle al di fuori del loro intervallo *T,* come le relative finestre.

Se si suppone che le frame siano tutte uguali (un *grano*), e le finestre *w* sono triangolari, dato che:

$$w(t)+w(t+T/2)+w(t+2\cdot T/2)+w(t+3\cdot T/2)+...=1$$

il segnale risultante non è altro che un "tappeto" costante ottenuto dalla sovrapposizione delle frame, privo di variazioni di ampiezza.

Vi sono molte importanti considerazioni, soprattutto sotto il profilo spettrale, da svolgere su una tecnica come questa, utilizzatissima nei contesti più disparati (sintesi a *wave-tables*, sintesi del parlato, sintesi additiva a FFT inversa, ecc.). Gli approfondimenti saranno oggetto di una trattazione a parte.

### *9.2 Il* **ringing,** *o fenomeno di Gibbs*

Si è già detto (pag. 5) che per esprimere nei termini dell'analisi armonica un segnale periodico qualunque può risultare necessario un numero infinito di armoniche. Questo è il caso di segnali anche elementari come i denti di sega, le onde quadre, triangolari e gli impulsi, a causa della circostanza che comprendono discontinuità o punti angolosi,

Un'onda quadra di frequenza $\omega_1$ è esprimibile come somma di armoniche di ampiezza decrescente con la legge $1/n$, e dove le armoniche pari siano siano soppresse:

$$\text{quadra}(t)=\sum_1^{\infty}\text{dispari}(n)\frac{1}{n}\cdot\cos(n\cdot\omega_1\cdot t)$$

dove $\text{dispari}(n)$ è un "selettore" che vale 1 se *n* è dispari e zero altrimenti, e quindi azzera le armoniche pari. La fase per questo tipo di segnale, se è sincronizzato sull'origine dei tempi, è zero.

Un dente di sega è esprimibile nello stesso modo, ma senza soppressione delle armoniche pari:

$$\text{dentedisega}(t)=\sum_1^{\infty}\frac{1}{n}\cdot\cos(n\cdot\omega_1\cdot t)$$

Un impulso (un treno di impulsi con frequenza $\omega_1$ ) infine è ottenuto (come abbiamo già avuto modo di vedere) dalla somma di armoniche tutte con ampiezza unitaria:

$$\text{impulso}(t)=\sum_1^{\infty}\cos(n\cdot\omega_1\cdot t)$$

Se osserviamo ora le espressioni della quadra e del dente di sega notiamo che il contributo nelle armoniche decresce con la loro altezza. Ci si deve quindi aspettare che il loro ruolo diventi via via più trascurabile; rigorosamente parlando, tuttavia, tutte le infinite armoniche contribuiscono al segnale risultante. Questo non è invece il caso del treno di impulsi, nel quale tutte la armoniche, indipendentemente dalla loro altezza, forniscono il medesimo contributo.

Sappiamo che in un sistema campionato nel tempo è possibile esprimere solo un intervallo limitato di frequenze (esattamente, fino alla frequenza di Nyquist, che è pari alla metà della frequenza di campionamento). Ci si domanda quindi cosa accada a questi segnali di sintesi quando si sia costretti, o per questo motivo, o per un altro qualsiasi, ad utilizzare un numero finito di frequenze. Dal punto di vista dell'ascolto, se le frequenze così soppresse sono oltre l'udibile, non ci aspettiamo – come già fatto notare – nessuna differenza. Ma dal punto di vista della forma d'onda, cosa succede?

Se proviamo a generare una forma d'onda (ad es. un dente di sega) con un numero finito di armoniche vediamo comparire un fenomeno, detto *fenomeno di Gibbs*, o anche *ringing*, che consiste in una oscillazione tanto più accentuata e a frequenza tanto più bassa quanto più bassa è la frequenza dalla quale inizia la soppressione (chiamiamola *frequenza di taglio*).

Questa oscillazione è simmetrica nel tempo e raggiunge il massimo nei punti dove la funzione “originale” eseguirebbe un brusco salto. Si deve in effetti notare che è proprio in questo punto che sarebbero necessarie le armoniche di frequenza alta (che sono state soppresse) per esprimere questo brusco salto. Ci si deve quindi aspettare che il fenomeno sia assente o molto meno presente per segnali privi di bruschi salti (discontinuità). La frequenza del fenomeno è pari alla massima frequenza presente nel segnale (quindi, alla frequenza di taglio).

Se utilizziamo un numero elevato di armoniche, il segnale ottenuto sarà molto poco differente da quello che si desiderava sintetizzare, al limite indistinguibile. Va infine fatto notare che nei sistemi numerici campionati il *ringing* è sempre presente, ma alla frequenza di Nyquist, che di norma non è udibile.

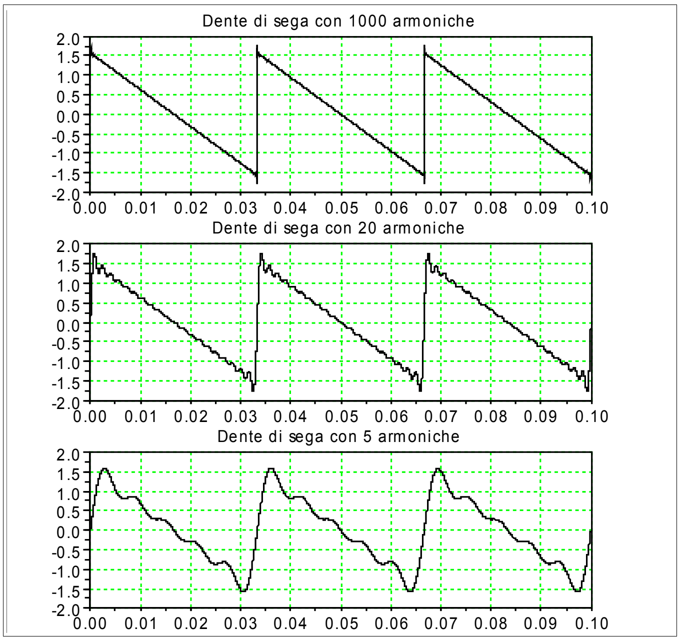


Fig. 12 - Dente di sega realizzato con diverse quantità di armoniche che mostra il *fenomeno di Gibbs (ringing)*

Abbiamo chiamato la frequenza superiore presente nella sintesi del segnale *frequenza di taglio*. In effetti il brusco troncamento della serie di seni può essere interpretato come un filtraggio passa-basso *infinitamente ripido* con frequenza di taglio pari a quella di troncamento. Filtri siffatti, con pendenza infinita (detti anche *filtri a muro*, con significativa metafora, o anche *filtri passa-basso ideali*) sono irrealizzabili nel continuo (quindi, nell'elettronica analogica) perché richiederebbero infiniti componenti, ma lo sono in quella numerica. Basti pensare, ad esempio, di fare un'analisi di Fourier del segnale in ingresso, di sopprimere tutte le frequenze superiori a quella voluta, e di risintetizzare il segnale: si farebbe qualcosa di analogo a quanto fatto qui sopra con il dente di sega.

Il fenomeno di Gibbs, o *ringing*[33], è quindi legato anche alla presenza di filtraggio (esplicito o implicito) di tipo ideale, con pendenze infinite o molto elevate.

Sull'onda di quanto affermato fin'ora, ritornando alla considerazione che il *ringing* è legato alla presenza nel segnale da sintetizzare di fronti ripidi (discontinuità) nel dominio del tempo, e di fronti ripidi (discontinuità) nel dominio della frequenza, un modo per mitigare il fenomeno è tentare di sintetizzare un segnale simile a quello di partenza, ma dotato di fronti meno ripidi, ovvero, nel dominio della frequenza, di rendere meno brusca la caduta di ampiezza nella frequenza. Questo può essere ottenuto applicando delle finestre "dolci" nel dominio della frequenza, ovvero degli inviluppi in frequenza, che rendano meno brusca la caduta a zero:

$$\mathrm{output}(t)=\sum_{1}^{N} a_n \cdot \mathrm{w}(n) \cdot \cos(\omega_n \cdot t+\phi_n) \qquad 18$$

dove $w(n)$ vale zero per $n \geq N$, $w(1)=1$, e fra i due valori scende con "dolcezza".

Esistono molte finestre (funzioni matematiche) studiate appositamente per questo genere di bisogna, ognuna dotata di sue specifiche proprietà. Una trattazione approfondita sarebbe qui fuori luogo, e ci si limita quindi a citare le più semplici. Anzitutto, la finestra triangolare.

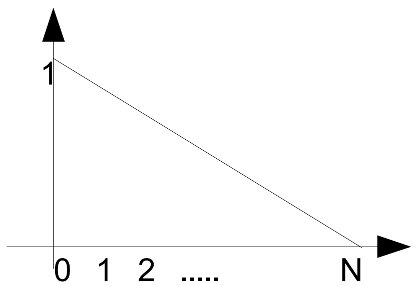

Fig. 13 - Finestra *w(n)* triangolare

Una finestra del genere ha la forma:

$$\mathrm{w}(n)=\frac{-1}{N}\cdot n+1 \quad \mathrm{per}\, n \leq N\,; \quad \mathrm{w}(n)=0 \quad \mathrm{per}\, n>N$$

Se si sintetizza il dente di sega applicando la finestra triangolare come indicato nella 18, si ottengono i risultati seguenti.

33 *Ring* indica in inglese il suono del campanello.

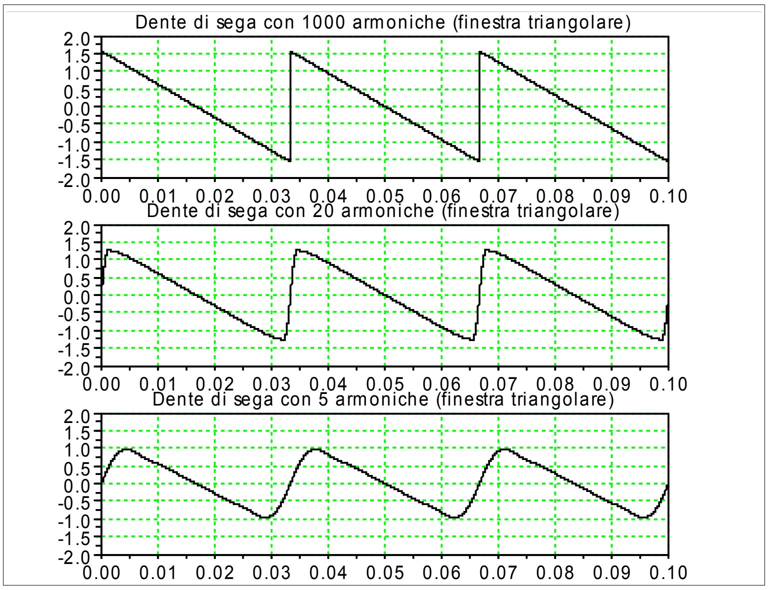


Fig. 14 - Stesso dente di sega, ma con finestra triangolare.

L'introduzione di una finestra rende meno visibile il *ringing* (che sarebbe udibile, se la frequenza di taglio fosse in banda audio), e "addolcisce" gli spigoli della forma d'onda, pur rispettandone le caratteristiche geometriche di fondo. Già con 5 armoniche la forma è accettabile.

Un'altra finestra molto utilizzata, meno brutale del triangolo, è il "coseno rialzato", che porta a risultati solo lievemente migliori:

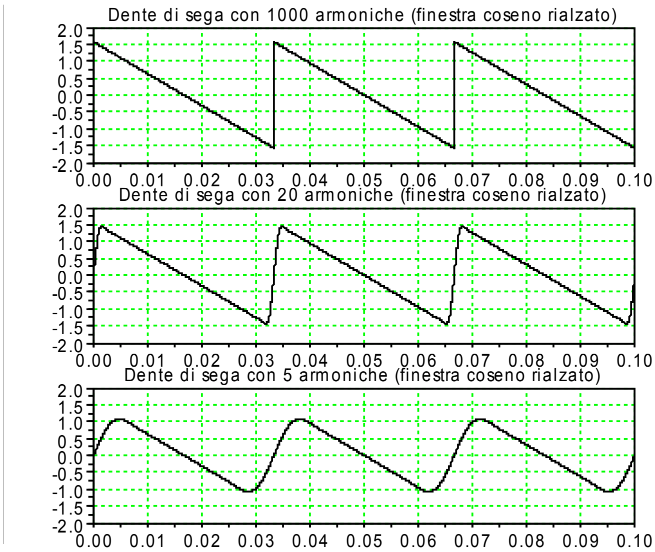


Fig. 15 - Dente di sega con finestra a coseno rialzato

La finestra "coseno rialzato" (*raised cosine*) è ottenuta sommando 1 ad un coseno e dimezzandone l'ampiezza (in pratica, "sollevando" e attenuando il coseno in modo che vari da 1 a 0). l'azzeramento delle frequenze è meno accentuato alle basse, e più rapido alle alte, rispetto al triangolo.

$$w(n)=\frac{1}{2}(\cos(\pi\frac{n}{N})+1) \;\text{ per }\; n\leq N\,;\; w(n)=0 \;\text{ per }\; n>N$$

E' piuttosto immediato notare come l'applicazione di una finestra (come quelle qui esemplificate) sia equivalente ad un filtraggio passa-basso (anche se irrealizzabile", o difficilmente realizzabile, nell'analogico), dato che si attenuano in maggior misura le frequenze alte. L'aspetto “non realizzabile” nell'analogico corrisponde al fatto che una finestra porta matematicamente a zero una frequenza precisa, mentre i filtri analogici possono sopprimere una banda di frequenze, ma hanno tutti un andamento asintotico (verso frequenze infinite) di tipo log-log lineare (una pendenza multipla di 6 dB/ottava) e, come detto, non permettono “pendenze infinite”.

http://www.mnt-aq.it
Versione: 0.4 del 28 settembre 2010
Testi, formule e figure: OpenOffice
Grafici, calcoli: Scilab
Calcolo simbolico: Maxima