QUANTIZZAZIONE DEI SEGNALI

Riprendiamo ora il caso del campionamento monodimensionale di un segnale reale f(t); esso, una volta campionato, è costituito da una successione di impulsi che si presentano ad intervalli di tempo di estensione $2\pi/\omega_c \leq \pi/\omega_0$.

La codifica di questi impulsi può essere effettuata mediante metodi analogici (esempi: Pulse Amplitude Modulation, Pulse Duration Modulation, Pulse Position Modulation; vedi figura nel lucido) o mediante metodi numerici.

La codifica numerica di questi impulsi si ottiene mediante quantizzazione nel discreto dei valori reali degli impulsi-campione (Pulse Code Modulation).

Quantizzazione lineare

Se operiamo i seguenti passi:
a) dividiamo la banda d'ampiezza del segnale f(t) in intervalli aventi larghezza fissa ∂,
b) associamo un'unica parola di codifica (intera) in corrispondenza con tutti i valori (reali) di un particolare intervallo,
c) ripetiamo l'operazione per tutti gli intervalli della banda di ampiezza,

allora otteniamo la **quantizzazione lineare** della banda d'ampiezza data.

Possiamo considerare, ad esempio, di associare una codifica binaria con parole di 3 bit per una quantizzazione a 8 intervalli della banda d'ampiezza mediante la parola binaria

$q = c_1 c_2 c_3$

che rappresenta il numero

$c_1 \times 2^2 + c_2 \times 2^1 + c_3 \times 2^0$

compreso tra 0 e 7, associato ad uno specifico intervallo ed in cui c_1, c_2, c_3 sono cifre binarie.

Nella figura del lucido possiamo vedere l'associazione tra il valore campionato e la sua codifica binaria a 3 bit.

Quando la codifica è realizzata mediante un impulso per le cifre 1 e nessun impulso per le cifre 0 viene detta PCM binaria unipolare.

Quando la codifica è realizzata mediante un impulso positivo per le cifre 1 e un impulso negativo per le cifre 0 viene detta PCM binaria polare.

Se i livelli degli impulsi fossero più di due (positivo/nullo o positivo/negativo) avremmo codifiche PCM N-arie dove N è il numero dei livelli per impulso.

Nel caso di PCM binario abbiamo 2^k livelli di quantizzazione con una parola di codifica costituita da k cifre binarie.

Nel caso di PCM N-ario abbiamo N^k livelli di quantizzazione con una parola di codifica costituita da k cifre del sistema di numerazione N-ario.

La codifica più diffusa è comunque binaria poiché è quella che consente una minore introduzione di rumore.

Ricordiamo che la codifica di un segnale sorgente campionato mediante quantizzazione produce sempre una certa degradazione

del segnale dovuta all'introduzione di rumore (ovvero di errore nella codifica del segnale).

Supponiamo, in generale, di avere un segnale continuo f(t), che il numero di intervalli di ampiezza sia M (intero, pari), f_k sia un valore tale che

$$f_k = k\partial$$

con

$$k = [-M/2,-(M/2)+1,...,-1, 0,1,2,...,(M/2)-1]$$

e ∂ l'ampiezza reale di ogni singolo intervallo;

allora l'applicazione Q definita sui reali e a valori nell'insieme

(1) $I_s = \{f_k\}_{k=-M/2,(M/2)-1}$

definita da

(2) $Q(f(t)) = f_k(t)$ tale che
$(f(t)-f_k(t)) = \min_{k=-M/2,(M/2)-1} (f(t)-f_k))$

è detta **quantizzazione** e l'errore

(3) $e_k(t) = (f(t) - f_k(t))$

derivante dalla approssimazione effettuata viene chiamato **rumore di quantizzazione** e assume valori nello intervallo $[0, \partial/2]$ se $|f| < \partial M/2$ e nell'intervallo $[0, +\infty]$ se $|f| \geq \partial M/2$.

Nel primo caso è detto **errore di granularità**, nel secondo **errore di sovraccarico.**

Se il segnale f(t) era un segnale campionato, allora il segnale $f_k(t)$ sarà un segnale campionato e quantizzato ovvero discreto sia rispetto al tempo che rispetto all'ampiezza.

Allora, poiché sarà $f_k(t)$ il segnale campionato e quantizzato che verrà trasmesso, il valore massimo del rapporto tra il segnale sorgente e il rumore di quantizzazione introdotto, cioè il rapporto ottimale, sarà

$$(4) \quad SNR = \overline{f(t)^2} \,/\, \overline{e_k(t)^2}$$

La valutazione di $\overline{e_k(t)^2}$ implica la media di $e_k(t)^2$ per tutta l'estensione temporale del segnale. Possiamo considerare che l'errore sia lineare per tutta l'estensione di ampiezza considerata (essendo lineare la suddivisione in M intervalli di ampiezza ∂ dell'ampiezza del segnale) e ottenere

$$(5) \quad \overline{e_k^2} = \int_{e_k=-\partial/2..+\partial/2} e_k^2 \, d(e_k/\partial) = \partial^2/12$$

che descrive l'**errore quadratico medio** del processo di quantizzazione. Possiamo notare dalla (5) come, per quanto piccolo sia ∂, il rumore di quantizzazione non sia mai nullo.

Considerazioni sugli intervalli di quantizzazione

Consideriamo il segnale f(t) come un segnale stocastico (random).

Se il segnale f(t) è **stazionario** (ovvero invariante dal punto di vista statistico rispetto a traslazioni temporali) è possibile ricavare il valore ottimale di ∂ per un prefissato numero di intervalli M dalla distribuzione di probabilità delle ampiezze di f.

Nel caso di segnali **non stazionari** è generalmente opportuno considerare soluzioni in cui il valore di ∂ sia adattivo, al

fine di minimizzare l'errore di quantizzazione. Per ∂ adattivo intendiamo che esso vari in dipendenza della banda di ampiezza della porzione di segnale da codificare che viene considerata.

Inoltre, quando il segnale sorgente varia in una banda di ampiezza larga, ovvero varia nel tempo in un intervallo che copre alcuni ordini di grandezza (segnali ad ampia gamma dinamica), se il valore di ∂ è fisso, al variare dell'energia del segnale varia il rapporto tra le componenti del rumore di quantizzazione prodotto.

Al diminuire dell'ampiezza del segnale sorgente il rumore di sovraccarico scompare mentre quello di granulazione aumenta fino a coincidere con il segnale quando questo diventa più piccolo di ∂.

Conviene allora considerare una delle seguenti due soluzioni per evitare che il rapporto segnale/rumore si abbassi troppo:
a) utilizzare una parola di codifica costituita da un numero di cifre sufficientemente grande da far sì che ∂ sia sempre molto più piccolo del segnale sorgente;
b) adattare ∂ in funzione dell'ampiezza del segnale sorgente.

La seconda soluzione è ovviamente quella che consente di ottenere una rappresentazione più efficiente.

Consideriamo ad esempio un generico segnale sonoro; esso deve essere certamente considerato un segnale non stazionario ad ampia gamma dinamica (tanto in alcuni suoni accidentali quanto negli attacchi di certi strumenti musicali vi sono variazioni di energia che lo dimostrano) anche se vi sono tuttavia casi particolari (il parlato ad esempio) in cui si può rilevare una

stazionarietà almeno "locale" cioè limitata ad un certo intervallo di tempo.

Compressione/espansione (companding)

Definiamo fattore di picco di un segnale f(t) il rapporto

(6) $K = \left(|f(t)|_{max}^2 / \overline{f^2(t)} \right)^{1/2}$

Il fattore di picco rappresenta un indicatore dell'efficienza della codifica poiché ad alti valori del fattore di picco l'efficienza è bassa e viceversa.

Infatti, basta considerare un segnale con valor medio nullo per osservare che il valore di ∂ dipende direttamente (e linearmente) dall'ampiezza di picco del segnale

(7) $\partial = 2 |f(t)|_{max} / 2^N = 2 |f(t)|_{max} / M$

con N numero di cifre binarie della parola di codifica e $M = 2^N$ il numero di livelli di quantizzazione.

Inoltre, applicando questo risultato nella (5) e nella (4) otteniamo

(8) $SNR = 3M^2 \overline{f^2(t)} / |f(t)|_{max}^2 = 3M^2/K^2$

Per ottenere una migliore efficienza della parola di codifica è stato definito il metodo di compressione/espansione (**companding**) che ha la funzione di migliorare il SNR per i valori medio-bassi del segnale (vedi figura nel lucido).

Lo stesso effetto potrebbe essere ottenuto con un quantizzatore non lineare o anche trascurando certi livelli (più vicini al valore di picco) di un quantizzatore lineare.

Quantizzazione floating point

Per i segnali sorgenti che sono non stazionari e ad ampia gamma dinamica la quantizzazione può essere effettuata codificando il valore di ampiezza del segnale con un codice numerico di lunghezza fissa integrato da un termine moltiplicativo (anch'esso di lunghezza fissa) che ne consente la più opportuna collocazione nella gamma dinamica; in altri termini, possiamo considerare questa codifica come costituita da una mantissa e un esponente; perciò si parla di quantizzazione a parola con virgola mobile o, più comunemente, **floating-point** o anche a intervallo di quantizzazione variabile.

L'errore di quantizzazione (di granularità) in questo modo aumenta o diminuisce coerentemente con la gamma dinamica del segnale sorgente (vedi figura nel lucido).

Poiché, inoltre, esistono diverse codifiche floating-point per il medesimo valore del segnale sorgente è evidente che non si tratta di un codice ottimale; d'altra parte, consente sia di "adattare" il rumore di quantizzazione alla dinamica del segnale che di comprimere la codifica lineare ottenibile con il metodo lineare di quantizzazione che abbiamo visto precedentemente.

Sono noti in letteratura principalmente tre diversi metodi per applicare la codifica a parola di quantizzazione floating-point:
a) **istantaneo:** per ogni numero campione del segnale viene calcolata la coppia mantissa-esponente; è ottimale come limitazione del rumore di quantizzazione, ma implica la codifica dell'esponente per ogni mantissa e perciò ha un fattore di compressione non ottimale;
b) **sillabico:** il valore dell'esponente varia adattivamente in funzione della gamma dinamica del segnale sorgente; quando l'ampiezza aumenta o diminuisce per un certo intervallo di tempo

viene variato l'esponente; richiede memoria proporzionalmente all'intervallo di tempo massimo considerato, ma consente di non codificare l'esponente se non quando viene variato; in tal caso dovrà ovviamente essere reso identificabile mediante un codice opportuno;
c) **a blocchi:** il valore dell'esponente viene codificato ogni N valori del segnale sorgente, perciò viene chiamata codifica floating-point a blocchi; richiede memoria per N valori di mantissa per poter calcolare l'esponente ottimale di ogni blocco.

Modulazione delta (DM)

La **modulazione delta** è un metodo di codifica del segnale campionato che si basa sull'approssimazione del valore istantaneo del segnale mediante somma o sottrazione di un "quanto" di ampiezza ∂.

Come possiamo osservare in figura (nel lucido), tranne il periodo transiente iniziale (in cui tutti i quanti sono positivi per raggiungere il valore del segnale partendo da 0) questo metodo consente di avere un errore che non supera mai $\pm\partial$, purché valga la condizione di avere una sufficiente frequenza di campionamento.

Una frequenza di campionamento insufficiente introdurrebbe altri transienti oltre a quello iniziale. Il rumore viene quindi espresso dalla seguente:

(9) $f_k(t) = f(t) + e_{Gk}(t) + e_{SOk}(t)$

dove $e_{Gk}(t)$ è la componente di rumore dovuta all'intervallo di modulazione delta ed $e_{SOk}(t)$ è la componente di rumore dovuta ai transienti in cui l'errore può eccedere $\pm\partial$; anche in questo caso, la prima viene detta errore di granularità e la seconda errore di sovraccarico (nei transienti).

Per comprendere più precisamente quali effetti comporta questo concetto consideriamo un segnale cosinusoidale

(10) $f(t) = A_f \cos(\omega_f t)$

La regione in cui il metodo di modulazione delta è soddisfacente corrisponde ai livelli

(11) $M/2 = A_f/\partial \leq (1/(2\pi))(\omega_s/\omega_f)$

poiché

(12) $\left|df(t)/dt\right|_{MAX} \leq \partial\omega_s/(2\pi)$

dove ω_s è la frequenza di campionamento.

Per esempio, nel caso $\omega_f/(2\pi) = 800$ Hz e $A_f/\partial = 20$ abbiamo

$f_s = \omega_s/(2\pi) \geq 20\omega_f = 40\pi(800) = 100{,}5$ kHz (approx)

ovvero abbiamo una frequenza di campionamento che è circa 62,8 volte la frequenza di Nyquist richiesta per un campionamento finalizzato alla quantizzazione lineare (cioè 1,6 kHz).

La modulazione delta ha trovato applicazione principalmente nella codifica di segnali vocali e di segnali video monocromatici.

Esistono varie estensioni del metodo di modulazione delta; di questi ricordiamo la modulazione delta adattiva (ADM) in cui il "quanto" di ampiezza ∂, su cui si basa il metodo, viene variato adattivamente quando si incontra un transiente, così da eliminare, o almeno limitare, la componente di errore di sovraccarico nei transienti. L'ADM trova sempre più applicazione nel campo audio.

Quantizzazione differenziale (DPCM)

La **quantizzazione differenziale** consiste nel codificare la differenza tra un valore campionato e il successivo invece dei valori stessi. A differenza della quantizzazione lineare, che essendo istantanea non richiede memoria, la quantizzazione differenziale richiede tanta memoria quanta necessaria per la codifica del campione precedente all'attuale.

La quantizzazione differenziale è simile alla modulazione delta, ma al posto di una coppia di valori (un bit) utilizza una parola di codifica costituita da un certo numero N di cifre binarie per ogni campione.

Un segnale sorgente f(t) campionato agli istanti di tempo $k\pi/\omega_0$ viene perciò codificato come un segnale

(13) $f'(k\pi/\omega_0) = f(k\pi/\omega_0) - f((k-1)\pi/\omega_0)$

Viceversa, il segnale f(t) verrà decodificato da $f'(k\pi/\omega_0)$ ottenendo i campioni corrispondenti agli istanti $k\pi/\omega_0$ secondo la seguente

(14) $f(k\pi/\omega_0) = f((k-1)\pi/\omega_0) + f'(k\pi/\omega_0)$

Nel lucido vediamo un esempio di segnale quantizzato in DPCM dove il treno di impulsi corrispondente alla codifica DPCM è costituito da parole di due bit per codificare i quattro intervalli differenziali utilizzati:

Ampiezza	Codifica
$+3\partial$	11
$+\partial$	10
$-\partial$	01
-3∂	00

La quantizzazione differenziale non introduce perdita di informazione frequenziale e consente una compressione della codifica tanto più vantaggiosa quanto più il segnale f(t) è caratterizzato da una banda in bassa frequenza, poiché questo corrisponde ad una derivata del segnale stesso caratterizzata da valori modesti e quindi a un intervallo più stretto di valori da codificare e da trasmettere.

Come nel caso della modulazione delta l'errore di quantizzazione è dato dalla seguente:

(15) $f_k(t) = f(t) + e_{Gk}(t) + e_{SOk}(t)$

e vale

(16) $\left|df(t)/dt\right|_{MAX} \leq (2^N - 1)\partial\omega_s/(2\pi)$

dove N è il numero di cifre binarie della parola di codifica.

Peraltro, senza conoscenza a priori sulla banda di frequenza del segnale sorgente non è possibile definire in modo ottimale le caratteristiche della parola di codifica e della frequenza di campionamento più opportuna. E' questo il limite principale della quantizza-zione differenziale.

Osserviamo a questo proposito che aumentando la frequenza di campionamento del segnale sorgente si può superare questo problema, a scapito ovviamente della compressione della codifica ovvero della quantità di codici da trasmettere, che torna così ad aumentare.

Aumentare la frequenza di campionamento può comunque servire anche a diminuire l'informazione da codificare per ogni singolo campione fino ad arrivare al caso limite in cui si codifichi un solo bit di informazione per codificare un campione differenziale (e si torna al caso della modulazione delta).

Nel caso della quantizzazione differenziale esistono in letteratura numerose varianti ed estensioni. Ricordiamo che anche in questo caso esiste ed è frequentemente applicato un metodo adattivo di quantizzazione differenziale (ADPCM).

Ad esempio, la codifica audio nei supporti ottici è di tipo PCM nei CD-DA, ma è ADPCM nei CD-I e nei CD-ROM XA (a vari livelli di qualità dipendenti da frequenza di campionamento e dimensione della parola di quantizza-zione differenziale).

Confronto tra i metodi di quantizzazione

La modulazione delta e la quantizzazione differenziale abbiamo già visto quanto siano strettamente legate; notiamo qui che confrontando il rapporto segnale/rumore (di granularità) dei due metodi abbiamo

(17) $SNR_{DPCM} = [(2^N - 1)^2 / (2N^3)] SNR_{DM}$

e perciò che il DPCM è superiore alla DM quando
(18) $N \geq 4$

Il confronto tra PCM e i metodi differenziali (DM, ADM, DPCM, ADPCM) non può prescindere dalla caratterizzazione del segnale da codificare.

Possiamo considerare in generale che in presenza di segnali stocastici la codifica PCM sia preferibile, mentre in presenza di segnali ben caratterizzati si possa individuare una codifica differenziale più efficiente del PCM.