Appunti di Teoria dei Segnali

Capitolo 13 - Trasmissione digitale

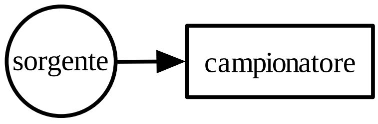
INTRODUZIONE

Fino ad ora, nello studio della codifica e della trasmissione numerica dei segnali, ci siamo occupati di sorgenti tempo-discrete a valori discreti, ossia di sorgenti che possiedono un alfabeto costituito da un numero finito (o infinito numerabile) di simboli, i quali vengono emessi ad intervalli di tempo regolari. Adesso generalizziamo ulteriormente i nostri concetti descrivendo, a livello molto più qualitativo che quantitativo, cosa si fa quando c'è da trasmettere, in forma digitale (ossia come sequenza di bit), un generico segnale analogico, ossia un segnale continuo nel tempo e a valori continui.

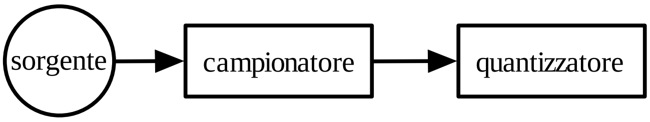
Se la trasmissione di questo segnale deve essere fatta in forma numerica, lo schema di trasmissione da considerare deve essere tale da permettere la codifica del segnale in forma binaria e, successivamente, la decodifica del segnale binario (al fine di far arrivare al ricevitore lo stesso segnale analogico che ha emesso la sorgente).

Quindi, il primo passaggio fondamentale da compiere sul segnale analogico di partenza s(t) è il suo **campionamento**: la funzione del campionamento, come sappiamo e come sarà ripetuto tra poco, è quella di generare i campioni del segnale che poi consentiranno di effettuare la ricostruzione (esatta solo a livello teorico) del segnale stesso.

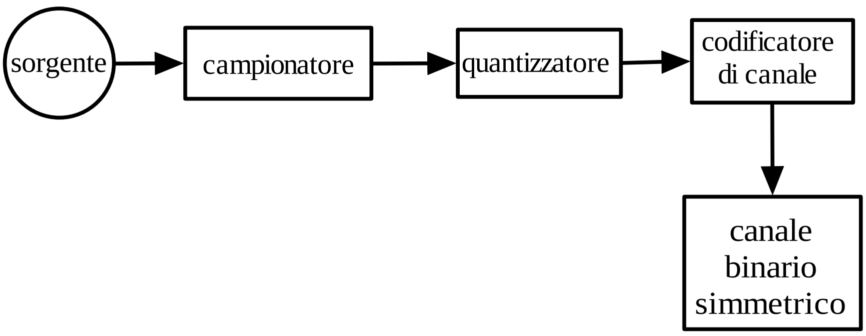
Possiamo quindi cominciare a tracciare lo schema del nostro sistema di trasmissione:



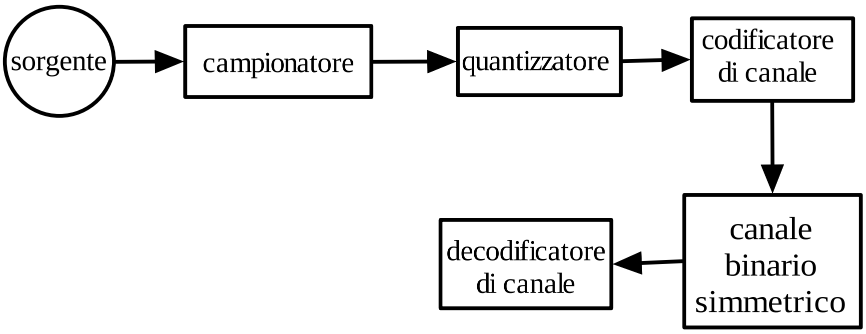
In uscita dal campionatore, abbiamo dunque i campioni del segnale di partenza e sono questi campioni che dobbiamo trasmettere in forma binaria: per farlo, dobbiamo effettuare una **quantizzazione**, ossia un procedimento che associa, ai possibili valori del segnale considerato, delle sequenze di bit.



In uscita dal quantizzatore abbiamo quindi una sequenza di bit, che dobbiamo trasmettere, mediante un **canale binario** (che supporremo per semplicità simmetrico), al ricevitore. Anche questo aspetto è stato già analizzato in precedenza e, in particolare, si è visto come, data la presenza inevitabile di errori dovuti alla trasmissione nel canale, è opportuno compiere delle manipolazioni, sulla sequenza di bit, che consentano, in fase di ricezione, di recuperare tali errori e di ottenere comunque la sequenza binaria corretta. L’opera di codifica del segnale binario da trasmettere è compiuta dal **codificatore di canale**: esso effettua le suddette manipolazioni sul segnale binario da trasmettere e invia tale segnale al canale, il quale si occupa della trasmissione vera e propria.



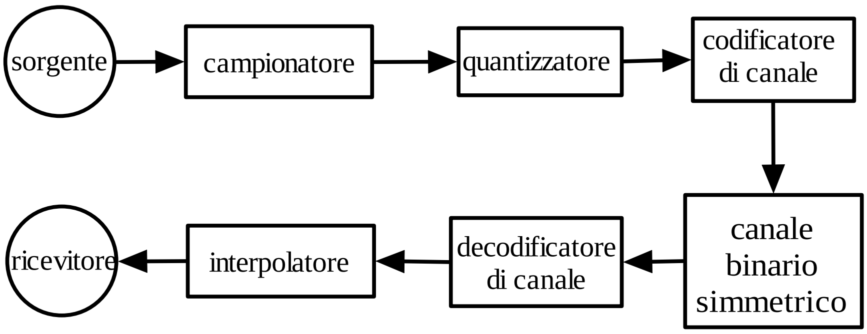
All’uscita dal canale binario si ottiene dunque una sequenza di bit verosimilmente affetta da errori: questi errori vengono recuperati (nel miglior modo possibile) da un altro dispositivo che prende il nome di **decodificatore di canale**, il quale si comporta in modo praticamente inverso al codificatore di canale.



Quello che viene fuori dal decodificatore di canale è dunque, in linea di massima, la stessa sequenza di bit che è stata generata dal quantizzatore, ossia il codice binario corrispondente ai campioni del segnale analogico. A partire da questa sequenza di bit, è necessario ricostruire il segnale analogico emesso dalla sorgente: a questo compito si dedica un organo che prende il nome di

interpolatore. In poche parole, questo organo si occupa semplicemente di applicare la formula di ricostruzione vista per il campionamento.

All'uscita dell'interpolatore otteniamo dunque il segnale <u>analogico</u> emesso dalla sorgente e questo segnale giunge infine al ricevitore, per cui possiamo completare il nostro schema nel modo seguente:



Dobbiamo adesso occuparci un po' più nel dettaglio di come funzionano, nella realtà, i dispositivi inclusi in questo schema.

CAMPIONAMENTO REALE

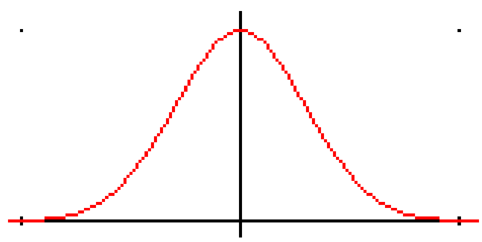
Del campionamento <u>ideale</u> si è parlato abbondantemente in precedenza, per cui facciamo un rapido riepilogo: intanto, "campionare" un segnale tempo-continuo s(t) significa determinare i valori s(nT) che il segnale in questione assume in istanti di tempo t=nT multipli di una quantità fissa T (quantità che prende il nome di "<u>*periodo di campionamento*</u>"). L'utilità del campionamento sta nella possibilità, sotto opportune ipotesi, di ricostruire con esattezza il segnale s(t) a partire dai suoi campioni s(nT). A livello teorico e sotto le ipotesi che la rendono possibile, la ricostruzione si effettua nel modo seguente: si considera intanto il cosiddetto "<u>*segnale campionato*</u>", ossia

$$s_C(t) = \sum_{n=-\infty}^{+\infty} s(nT)\delta(t - nT)$$

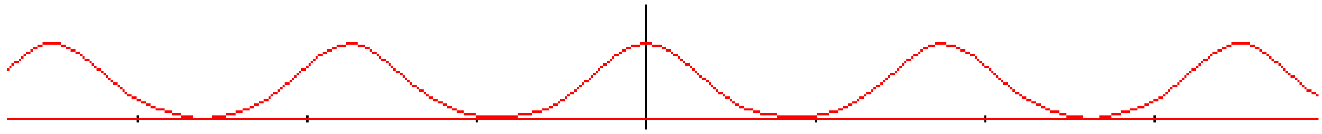
Questo è un segnale composto da una successione di impulsi equispaziati di T, ciascuno di area pari al valore assunto dal segnale nell'istante in cui è applicato l'impulso stesso. Si tratta evidentemente di un segnale tempo-continuo, per cui se ne può calcolare la trasformata di Fourier, che è

$$S_C(f) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} S\left(f - \frac{n}{T}\right)$$

Questa trasformata risulta composta da una successione di infinite repliche, a meno del fattore 1/T, dello spettro S(f) del segnale s(t). Per esempio, se lo spettro di s(t) è del tipo



lo spettro del segnale campionato è qualcosa del tipo



Allora si isola, da tale spettro, la componente centrata nell'origine, semplicemente moltiplicando $S_C(f)$ per un rettangolo esteso sulla banda di interesse:

$$S_r(f) = S_C(f)\left[\mathrm{Trect}\left(\frac{f}{f_C}\right)\right]$$

dove f_C=1/T è la "*frequenza di campionamento*" ed è stata presa in questo caso pari al doppio della banda del segnale s(t) (ossia, nel nostro caso, $f_c=2f_0$).

Infine antitrasformando $S_r(f)$, si ottiene il segnale s(t) ricostruito secondo la formula

$$\boxed{s_r(t) = \sum_{n=-\infty}^{+\infty} s(nT)\mathrm{sinc}\left(f_C(t-nT)\right)}$$

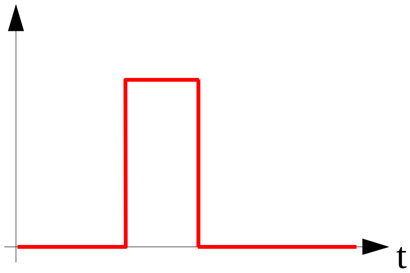
Il campionamento e la ricostruzione (o **interpolazione**) appena descritti sono quelli "ideali", dove l'idealità deriva essenzialmente da due fattori:

- il primo è la possibilità di prelevare, con esattezza, il valore che il segnale assume negli istanti di tempo considerati;

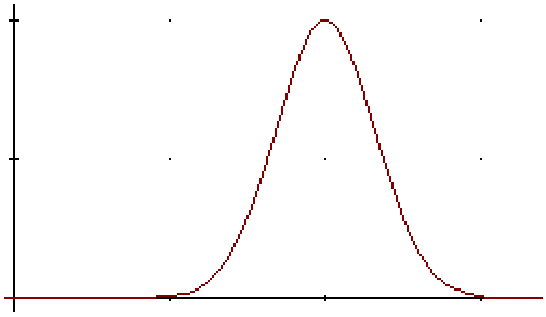
- il secondo è nella possibilità di ricostruire il segnale di partenza mediante una formula che utilizza un numero infinito di termini (cioè un numero infinito di campioni).

L'idealità è evidente, in quanto i dispositivi che noi utilizziamo nella realtà per effettuare il campionamento e l'interpolazione del segnale sono soggetti a limiti fisici che non consentono di effettuare le stesse operazioni previste a livello teorico. In particolare, le differenze fondamentali tra l'applicazione teorica del campionamento e quella pratica sono le seguenti:

- in primo luogo, nella pratica non è possibile prelevare il valore istantaneo assunto dal segnale considerato, mentre quello che si riesce a fare è soltanto prelevare i valori che il segnale assume nell'intorno degli istanti considerati; "prelevare" i valori che il segnale assume nell'intorno di un certo istante significa, a livello teorico, moltiplicare il segnale stesso per un impulso rettangolare, di altezza unitaria, centrato nell'istante considerato: l'effetto di questa moltiplicazione è quello di azzerare il segnale nell'intervallo in cui l'impulso è nullo e di isolare quindi il valore assunto dal segnale nell'intervallo in cui invece l'impulso vale 1; ovviamente, non è possibile ottenere concretamente un impulso rettangolare del tipo



ma quello che si riesce ad ottenere è qualcosa del tipo



da cui si intuisce l'approssimazione cui inevitabilmente si va incontro.

- in secondo luogo, non è ovviamente possibile prelevare infiniti campioni del segnale e poi ricostruire quest'ultimo utilizzando tali campioni; al contrario, è possibile prelevare solo un numero finito di campioni ed è con questi che va effettuata la ricostruzione (compiendo ovviamente anche qui delle approssimazioni).

Ad ogni modo, vediamo quali sono le condizioni sotto le quali è possibile ricostruire, in modo più o meno preciso, tale segnale.

Indicato con s(t) il segnale analogico che dobbiamo campionare, abbiamo detto che, per effettuare questo campionamento, dobbiamo moltiplicare s(t) per un altro segnale, che indichiamo con r(t), costituito da una successione di impulsi unitari centrati negli istanti in cui vogliamo prelevare i campioni. Il segnale r(t) prende il nome di **pettine di campionamento**.

In tal modo, il segnale campionato che otteniamo è del tipo $s_C(t) = s(t)r(t)$.

Il segnale r(t) è una successione di impulsi tutti uguali, centrati in istanti multipli di una quantità fissa nT: di conseguenza, è chiaramente un segnale periodico e come tale noi possiamo esprimerlo, a livello teorico, mediante il suo sviluppo in serie di Fourier, che sarà del tipo

$$r(t) = \sum_{k=-\infty}^{+\infty} r_k e^{j2\pi f_k t}$$

dove ovviamente i coefficienti dello sviluppo sono

$$r_k = \frac{1}{T}\int_{-T/2}^{T/2} r(t)e^{-j2\pi f_k t}dt$$

D'altra parte, r(t) è anche un segnale reale, per cui sappiamo di poterlo esprimere in modo più comodo come

$$r(t) = r_0 + \sum_{k=+1}^{+\infty} r_k \cos(2\pi f_k t)$$

Andando a sostituire nell'espressione del segnale campionato, abbiamo perciò che

$$s_C(t) = r_0 s(t) + s(t)\sum_{k=+1}^{+\infty} r_k \cos(2\pi f_k t)$$

Sviluppando parzialmente quella sommatoria, abbiamo che

$$s_C(t) = r_0 s(t) + s(t) r_1 \cos(2\pi f_1 t) + s(t) r_2 \cos(2\pi f_2 t) + \ldots$$

Passando al dominio della frequenza, ossia trasformando secondo Fourier ambo i membri di quest'ultima relazione, abbiamo che

$$S_C(f) = r_0 S(f) + r_1 S(f) * \text{Fourier}\left[\cos(2\pi f_1 t)\right] + r_2 S(f) * \text{Fourier}\left[\cos(2\pi f_2 t)\right] + \ldots$$

Ricordando adesso che la trasformata di Fourier della funzione Coseno è una somma di impulsi, abbiamo che

$$S_C(f) = r_0 S(f) + r_1 S(f) * \left[\frac{1}{2}\delta(f - f_1) + \frac{1}{2}\delta(f + f_1)\right] + r_2 S(f) * \left[\frac{1}{2}\delta(f - f_2) + \frac{1}{2}\delta(f + f_2)\right] + \ldots$$

Ricordando inoltre che la convoluzione tra una funzione qualsiasi e l'impulso di Dirac è pari alla funzione stessa calcolata nel punto di applicazione dell'impulso, abbiamo che

$$S_C(f) = r_0 S(f) + \frac{1}{2} r_1 S(f_1) + \frac{1}{2} r_1 S(-f_1) + \frac{1}{2} r_2 S(f_2) + \frac{1}{2} r_2 S(-_2) + \ldots$$

Questa formula mette in evidenza come, ancora una volta, lo spettro del segnale campionato sia costituito da una successione di repliche, a meno dei vari coefficienti, dello spettro di s(t); di tali repliche, una è centrata nell'origine (e scalata di una quantità r_0), per cui è proprio lo spettro di s(t), mentre le altre sono traslate rispetto ad essa. Di conseguenza, è possibile isolare la replica centrale, e quindi risalire, per antitrasformazione, ad s(t), solo a patto che la replica $r_0 S(f)$ non sia sovrapposta alle altre.

La conclusione è dunque che, anche nel campionamento reale, devono essere rispettate le condizioni di campionamento: questo significa che <u>lo spettro del segnale s(t) da ricostruire deve essere a banda limitata e significa anche che la frequenza di campionamento utilizzata deve essere almeno doppia della banda stessa di s(t)</u>.

Facciamo osservare una cosa importante: il fatto di esprimere r(t) mediante la formula

$$r(t) = r_0 + \sum_{k=+1}^{+\infty} r_k \cos(2\pi f_k t)$$

è possibile quale che sia r(t), a patto ovviamente che sia periodico e reale. Quindi, a prescindere da come è fatto nella realtà r(t), ossia in definitiva da come sono fatti gli impulsi che si riesce ad ottenere, quello sviluppo è comunque valido, per cui è valida anche nella realtà la formula di ricostruzione. Ovviamente, i problemi vengono dalla determinazione dei coefficienti dello sviluppo, in quali dipendono proprio dalle caratteristiche di r(t).

QUANTIZZAZIONE

Una volta generati i campioni del segnale, questi devono essere codificati in forma binaria e ciò avviene con il metodo della "quantizzazione", che consiste nei seguenti passi fondamentali:

- per prima cosa si individua il **range** del segnale, ossia l'intervallo $[s_{MIN}, s_{MAX}]$ nel quale rientrano tutti i valori assunti dal segnale stesso;

- tale intervallo viene successivamente diviso in un certo numero di intervallini più piccoli, i quali possono avere la stessa ampiezza (nel qual caso si parla di "*quantizzazione uniforme*") oppure ampiezza variabile (nel qual caso si parla di "*quantizzazione non uniforme*");

- a ciascuno di questi intervalli viene fatta corrispondere una sequenza di bit: per ogni intervallo, il numero N di bit associati è sempre lo stesso, il che significa che il numero di intervallini in cui viene diviso il range del segnale è 2^N (solitamente N=8);

- a questo punto, ogni campione del segnale rientrerà in uno di questi intervallini, per cui ad esso sarà associata la corrispondente sequenza di bit.

Questo è grossomodo il funzionamento del "quantizzatore": esso genera una sequenza di bit la quale viene trasmessa mediante il canale e quindi arriva all' "interpolatore"; quest'ultimo, come detto, ha in definitiva il compito di applicare la formula di ricostruzione vista a proposito del campionamento.

INTERPOLAZIONE

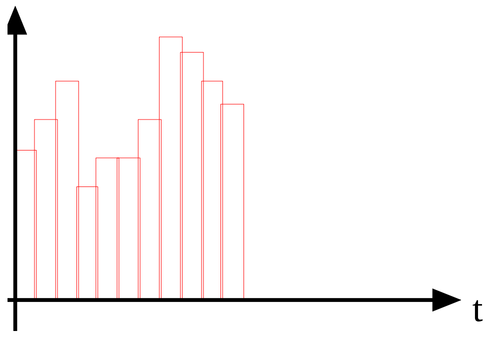
Dobbiamo adesso vedere come funziona nella realtà la fase di ricostruzione o "interpolazione" del segnale di partenza: a livello teorico, abbiamo ricordato che la ricostruzione avviene mediante la formula

$$s_r(t) = \sum_{n=-\infty}^{+\infty} s(nT)\mathrm{sinc}\left(f_C(t-nT)\right)$$

per cui si tratta di vedere, nella realtà, come applicare tale formula.

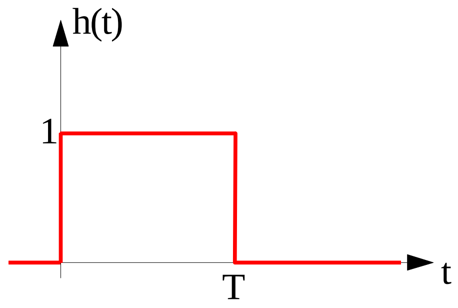
A questo proposito, il problema di fondo è il seguente: i campioni che arrivano, in forma binaria, all'interpolatore, non arrivano in modo continuo, ma intervallati tutti di uno stesso intervallo di tempo (comunque abbastanza piccolo); tali campioni, per come sono stati ottenuti, costituiscono i valori che il segnale di partenza assume nell'intorno di istanti predefiniti; si pone perciò il problema di quanto valga il segnale da ricostruire (che è un segnale continuo del tempo) nell'intervallo compreso tra l'arrivo di un campione e l'arrivo del campione successivo. Il modo più semplice di procedere è assumere che il segnale si mantenga costante in tale intervallo: per esempio, se, ad un certo istante arriva un campione che dà il valore 5, l'interpolatore assume che il segnale valga 5 finche non arriva il campione successivo.

L'effetto di un metodo di questo tipo è quello per cui il segnale viene approssimato mediante delle linee spezzate che vanno ciascuna da un campione al campione successivo:



Il vantaggio di un metodo di questo tipo è duplice:

- in primo luogo, quando i campioni sono molto ravvicinati nel tempo, l'approssimazione che si compie è davvero ottima;
- in secondo luogo, per realizzare un metodo di questo tipo è possibile utilizzare un sistema lineare tempo-invariante: tale sistema ha come risposta impulsiva la funzione



Infatti, indicato con x(t) il segnale che entra in ingresso, in un certo istante, a tale sistema e indicato con y(t) il corrispondente segnale in uscita, ci ricordiamo che essi sono legati mediante la relazione $y(t) = x(t) * h(t)$. Ma il segnale in ingresso x(t) non è altro che un impulso di area pari al valore assunto dal segnale in quell'istante, per cui abbiamo che

$$y(t) = A\delta(t) * h(t) = Ah(t)$$

da cui si deduce il motivo per cui la struttura del segnale è fatta di una successione di impulsi rettangolari.

Il dispositivo che effettua una ricostruzione del segnale tramite il metodo appena descritto prende il nome di **interpolatore di ordine 0** anche *mantenitore*.

Trasmissione numerica del segnale telefonico

INTRODUZIONE

Nei paragrafi precedenti abbiamo esposto a grandi linee come vengono realizzate nella pratica le operazioni di campionamento, quantizzazione e interpolazione di un segnale analogico. In questo, ci siamo riferiti ad un segnale analogico s(t) generico: adesso vogliamo scendere nei dettagli di ciò che si fa quando questo segnale è il **segnale telefonico**.

CAMPIONAMENTO

Abbiamo detto che i requisiti di fondo perché il campionamento reale (come anche quello ideale) possa consentire la ricostruzione del segnale di partenza a partire dai sui campioni, sono essenzialmente due:

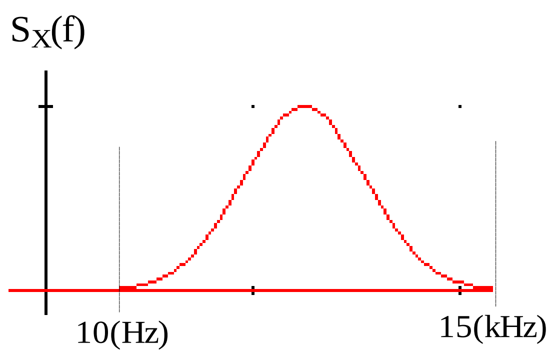
- in primo luogo, il segnale in esame deve essere "a banda limitata", ossia deve avere una trasformata di Fourier che assume valori sempre nulli al di fuori di un certo intervallo;
- in secondo luogo, la frequenza di campionamento deve essere presa almeno doppia della banda del segnale in esame (teorema del campionamento).

E' evidente, allora, che, per progettare i dispositivi che devono effettuare il campionamento, noi dobbiamo conoscere le caratteristiche spettrali del segnale da campionare.

Vediamo allora qual è il segnale che noi dobbiamo campionare. Quando noi parliamo (a prescindere dal fatto che lo facciamo davanti alla cornetta telefonica o meno), emettiamo un segnale che prende evidentemente il nome di **segnale vocale**; in linea di massima, è questo segnale che noi dobbiamo trasmettere da un capo all'altro del sistema di trasmissione. Il segnale vocale non è chiaramente un *segnale deterministico*, nel senso che noi, nel progettare un sistema di trasmissione, dovendo metterci nelle condizioni di poter trasmettere un segnale dalle caratteristiche più disparate, non facciamo i nostri ragionamenti su un segnale preciso, ma su un segnale assolutamente generico. E' chiaro allora che consideriamo il segnale vocale come un *segnale aleatorio*, per cui lo studiamo in termini statistici.

In particolare, studiando lo spettro (che in questo caso è lo spettro di potenza) di tale segnale, si è trovato che esso è a banda limitata e, soprattutto, comprende frequenze tra **10 Hz** e **15000 Hz**.

Supponiamo, ad esempio, che si tratti di uno spettro del tipo seguente:



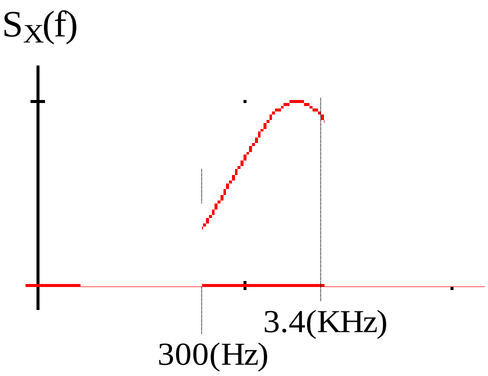
Il fatto che il segnale vocale sia a banda limitata (di circa **15 kHz**) ci indica subito che possiamo campionarlo e poi ricostruirlo mediante tutti i ragionamenti fatti prima. Il passo successivo, per cui la ricostruzione possa avere successo, è la scelta della frequenza di campionamento, che, ricordiamo ancora una volta, deve essere almeno il doppio della banda del segnale da campionare.

Per effettuare la scelta di tale frequenza ci si pone la seguente domanda: dovendo trasmettere il segnale vocale, è davvero necessario trasmetterlo "interamente", ossia è davvero necessario, al fine di ottenere una buona qualità del servizio, trasmettere tutta la potenza associata al segnale, oppure è possibile inviarne solo una parte?

La risposta a questa domanda dipende, come sempre, dalle circostanze:

- *nei sistemi ad alta fedeltà*, è necessario raggiungere una qualità della trasmissione particolarmente alta, per cui deve essere trasmessa tutta (o quasi) la potenza associata al segnale; ciò significa che la frequenza di campionamento deve essere effettivamente pari almeno al doppio della banda del segnale vocale: essendo tale banda di circa 15 kHz, la frequenza di campionamento nei sistemi ad alta fedeltà è superiore ai 30 kHz; tanto per fissare le idee, la frequenza di campionamento usata per registrare i file musicali (in cui c'è la somma di segnale vocale e suoni prodotti dagli strumenti) è di **48 kHz**, ma ci sono casi in cui la frequenza di campionamento scelta è anche maggiore;

- viceversa, *nei comuni sistemi di trasmissione telefonica*, si è stimato che è possibile ottenere una più che accettabile qualità del servizio trasmettendo anche una piccola parte della potenza del segnale vocale[1]; questo significa che, nella trasmissione telefonica, non sia il segnale vocale ad essere trasmesso, ma solo parte di esso; si dice perciò che si effettua un **filtraggio** del segnale vocale, ossia lo si priva delle componenti di frequenza ritenute inutili. In particolare, è stato stimato che l'intervallo di frequenza sufficiente per la trasmissione è $\left[300(Hz), 3400(Hz)\right]$, per cui il filtraggio consiste nell'eliminare le componenti di frequenza esterne a tale intervallo. Con riferimento al diagramma in frequenza disegnato prima, l'esito del filtraggio fornisce il seguente spettro:



Il segnale che ne viene fuori è il cosiddetto **segnale telefonico**[2] ed è quello che effettivamente viene trasmesso: chiaramente, quindi, essendo pari a 3400(Hz) la banda di questo segnale (misurata dalla frequenza f=0), la frequenza minima di campionamento da utilizzare deve essere 6800(Hz); nella pratica, viene preso il valore di **8000(Hz)** (scelto in base alle caratteristiche fisiche degli apparati usati per la trasmissione).

[1] Infatti, in un sistema telefonico, quello che conta principalmente, durante una conversazione, è che ogni utente possa fare solo due cose: riconoscere il proprio interlocutore e capire quello che dice. Si è allora verificato che, per ottenere questi due risultati, è sufficiente trasmettere solo una minima parte della potenza del normale segnale vocale.

[2] detto anche **segnale vocale di qualità telefonica**.

Quindi, la frequenza utilizzata per il campionamento del segnale telefonico è $f_C = 8000(\text{Hz})$, il che significa che vengono prelevati 8000 campioni del segnale ogni secondo. Ovviamente, essendo il periodo di campionamento pari all'inverso della frequenza di campionamento, ciò che si trova è che

$$T_C = \frac{1}{f_C} = 125(\mu\,\text{sec})$$

Quindi, per la trasmissione numerica del segnale telefonico usiamo dispositivi che prelevano campioni del segnale telefonico ogni **125 μsec**.

QUANTIZZAZIONE

I campioni così prodotti vanno poi quantizzati. Come sappiamo, il problema più grosso della quantizzazione sta nel ridurre quanto più è possibile il cosiddetto **rapporto segnale-rumore**, ossia il rapporto tra la potenza associata al segnale quantizzato e la potenza associata al **rumore di quantizzazione** ad esso sovrapposto. A suo tempo abbiamo visto che questo rapporto è strettamente legato al numero di bit usati per la quantizzazione, ossia quindi al numero di *intervallini* in cui viene diviso il range dinamico del segnale.

Quando abbiamo studiato la quantizzazione, abbiamo preso in esame la cosiddetta **quantizzazione uniforme**, caratterizzata dal fatto che l'ampiezza dei suddetti intervallini è sempre la stessa. Ora, una tecnica di questo tipo è consigliabile quando il segnale da quantizzare ha una *distribuzione uniforme* all'interno del proprio range dinamico. Viceversa, quando ciò non accade, come per esempio per il segnale telefonico, questa tecnica è senz'altro meno vantaggiosa di quella della **quantizzazione non uniforme**, ossia di quella in cui l'ampiezza degli intervallini è maggiore da una parte e minore da un'altra.

Naturalmente, dove l'ampiezza dovrà essere minore (ossia ci sarà un maggiore infittimento degli intervallini) e dove sarà maggiore? E' chiaro che *dovremo prendere più intervallini in corrispondenza dei valori più probabili del segnale, mentre invece ci accontenteremo di meno intervallini (ossia di una peggiore approssimazione) in corrispondenza dei meno probabili.*

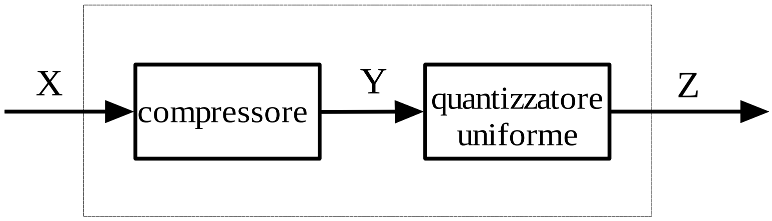
Questo discorso pone allora un problema essenziale: come si fanno a conoscere i valori più probabili e meno probabili assunti dal segnale telefonico? Ossia, in altre parole, come si fa a conoscere la distribuzione di tale segnale? E' molto difficile rispondere a questa domanda, per un motivo molto semplice: dato che noi facciamo i nostri ragionamenti non per un segnale specifico, deterministico, ma per un segnale aleatorio, possiamo considerare il segnale telefonico come un processo stocastico, che può dare cioè origine a infinite realizzazioni, ossia a infinite funzioni del tempo; per individuare le caratteristiche statistiche di tale processo, possiamo far uso del concetto di variabile aleatoria estratta: il problema sorge dal fatto che tale variabile aleatoria estratta non necessariamente ha sempre le stesse caratteristiche, ossia la stessa distribuzione; questo accadrebbe solo se il processo stocastico fosse stazionario. Al contrario, *il segnale telefonico è un processo stocastico non stazionario*, il che impedisce quindi di individuare una precisa distribuzione e quindi impedisce di impostare un unico metodo di quantizzazione.

L'altro problema fondamentale, come detto, è quello di massimizzare il rapporto segnale-rumore in uscita dal quantizzatore: sia nella quantizzazione uniforme sia in quella non uniforme, questo obbiettivo è perseguibile aumentando il numero di bit usati per la quantizzazione. Tuttavia, sappiamo bene che non sempre è questa la strada migliore, in quanto aumentare il numero di bit significa peggiorare i tempi di trasmissione e quindi la qualità del servizio. Al contrario, si verifica che proprio una opportuna quantizzazione non uniforme contribuisce ad aumentare il rapporto segnale-rumore, pur conservando sempre lo stesso numero di bit della quantizzazione uniforme.

In generale, quindi, diciamo che il quantizzatore va progettato studiando quanto più a fondo è possibile le caratteristiche statistiche del segnale telefonico e addattandolo ad esse. Questo però, come detto, è un discorso molto più complesso di quanto sembra, proprio in accordo al fatto che tali caratteristiche statistiche non sono costanti nel tempo.

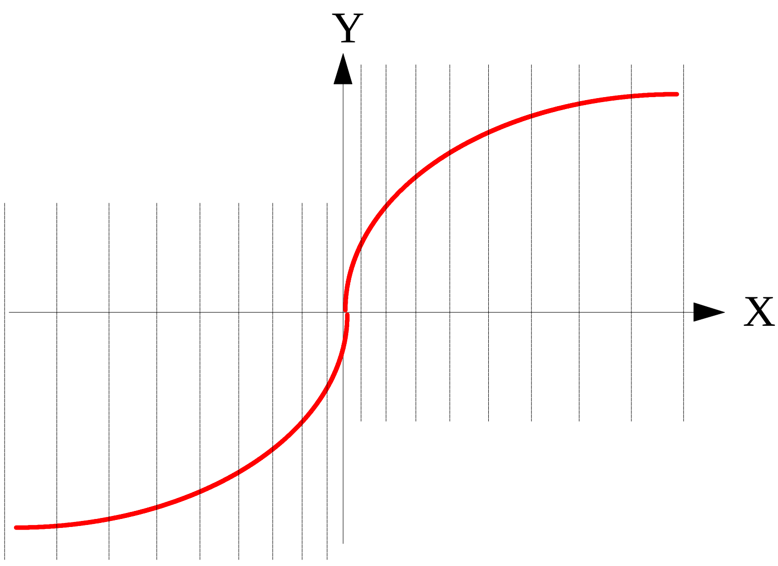
Struttura del quantizzatore

Premesso dunque il fatto che, per il segnale telefonico, viene adottata una quantizzazione non uniforme, in concreto, il quantizzatore che viene utilizzato è costituito da due diversi organi, che sono rappresentati nello schema seguente:



Il segnale X che arriva in ingresso al compressore è l'insieme dei campioni provenienti dal campionatore. Come sono fatti questi campioni? E' abbastanza realistico assumere che il segnale telefonico sia un segnale aleatorio a media nulla e sia perciò maggiormente distribuito intorno al valore 0; in altre parole, è lecito assumere che i valori più probabili di tale segnale siano quelli in prossimità dello zero, mentre quelli lontani dallo zero siano meno probabili.

Di questo fatto bisogna tenere conto nell'effettuare la quantizzazione non uniforme, nel senso che dobbiamo prevedere un maggiore infittimento degli intervallini in corrispondenza dei valori prossimi allo 0 e un minore infittimento in corrispondenza degli altri. A questo serve proprio il **compressore**, per il quale il legame tra il segnale in ingresso ed il segnale in uscita può essere schematizzato dal grafico seguente:



L'effetto di questo compressore è il seguente: operando una suddivisione del range dinamico di X tale che ci sia un maggiore infittimento degli intervallini in prossimità del valore 0, *il compressore fornisce una suddivisione praticamente uniforme del segnale in uscita Y.*

Tale segnale Y risulta dunque avere una distribuzione praticamente uniforme nel proprio range, per cui esso viene successivamente quantizzato, questa volta in maniera uniforme.

In uscita dal **quantizzatore uniforme** troviamo dunque il segnale Z che è il segnale binario prodotto appunto dalla quantizzazione.

La curva prima disegnata che lega i segnali X ed Y in ingresso ed in uscita al compressore, prende il nome di **caratteristica di compressione**; essa ha una duplice rappresentazione analitica, a seconda di quale sia lo standard utilizzato per la comunicazione telefonica:

- per il sistema telefonico numerico usato negli STATI UNITI, la legge matematica che rappresenta la caratteristica di compressione prende il nome di **μ-LAW** in base al fatto che in essa compare un parametro μ; si tratta della legge seguente:

$$Y = Y_{MAX}\frac{\ln\left[1+\mu\left(\frac{|X|}{X_{MAX}}\right)\right]}{\ln(1+\mu)}\mathrm{sgn}(X)$$

- per il sistema telefonico numerico usato invece in EUROPA, la legge matematica prende il nome di **A-LAW** (in base al fatto che in essa compare un parametro A) e si tratta della legge seguente:

$$Y = \begin{cases} Y_{MAX}\dfrac{A\left(\dfrac{|X|}{X_{MAX}}\right)}{1+\ln A}\mathrm{sgn}(X) & 0 \le \dfrac{|X|}{X_{MAX}} \le \dfrac{1}{A} \\[2ex] Y_{MAX}\dfrac{1+\ln\left[A\left(\dfrac{|X|}{X_{MAX}}\right)\right]}{1+\ln A}\mathrm{sgn}(X) & \dfrac{1}{A} \le \dfrac{|X|}{X_{MAX}} \le 1 \end{cases}$$

Si tratta di due equazioni che approssimano la curva di compressione prima visualizzata.

Per quanto riguarda nel dettaglio il funzionamento del quantizzatore uniforme, esso opera la quantizzazione servendosi di **8 bit**, ossia quindi dividendo il range dinamico del segnale Y in 256 intervallini tutti di uguale ampiezza. In tal modo, è stato stimato quanto vale il rapporto segnale-rumore sia per la μ-LAW sia per la A-LAW: tale rapporto assume in entrambi i casi il valore di **38 dB**, a patto che μ=255 e che A=87.56.

LA TECNICA PCM

Abbiamo dunque detto che lo standard internazionale prevede una frequenza di campionamento di 8000(Hz) ed una quantizzazione effettuata con 8 bit. Questi valori indicano, in pratica, che il segnale binario in uscita al quantizzatore, ossia il segnale che deve essere trasmesso (a meno di ulteriori bit di controllo aggiunti dal codificatore di canale) attraverso il canale, consiste di 64000 bit emessi al secondo: infatti, una frequenza di 8000(Hz) implica che vengano prelevati 8000 campioni al secondo e, se a ciascun campione vengono associati 8 bit, si ha che la "sorgente complessiva", ossia l'insieme di sorgente-campionatore-quantizzatore, emette un segnale composto da **64000 bit al secondo**.

Di conseguenza, il canale usato per la trasmissione deve avere una "capacità" di almeno 64000 bit al secondo, ossia deve poter trasmettere almeno 64000 bit al secondo.

L'insieme delle tecniche e degli standard finora descritti prende il nome di "**tecnica PCM**" dove questo acronimo sta per "**Pulse Code Modulation**".

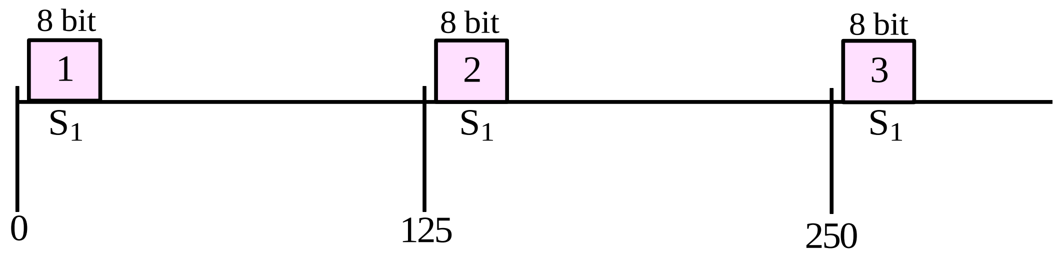
La quantità di 64 kbit/sec, proprio per i discorsi appena fatti, è di cruciale importanza nei sistemi di trasmissione numerici in cui anche il segnale vocale transita sui mezzi trasmissivi. Per questo motivo, essa è diventata una specie di unità di misura della capacità delle linee di trasmissione. Ad esempio, nella moderna **rete ISDN** (*Integrated Service Data Network*), si dice che un utente, stipulando un apposito contratto con il gestore telefonico (la Telecom Italia per intenderci), ha a disposizione (nel contratto base) una **linea ISDN** che corrisponde a due linee da 64 kbit/sec più una terza linea a 16 kbit/sec. Questo significa che l'utente può effettuare un collegamento ad Internet a 64 kbit/sec e, contemporaneamente, avere una linea libera, sempre da 64 kbit/sec, per parlare al telefono. La linea da 16 kbit/sec, invece, detta **linea di segnalazione**, non serve all'utente, ma alla rete telefonica per far funzionare il tutto (vi transitano tutte le informazioni di controllo e sincronizzazione).

LA TECNICA TDM

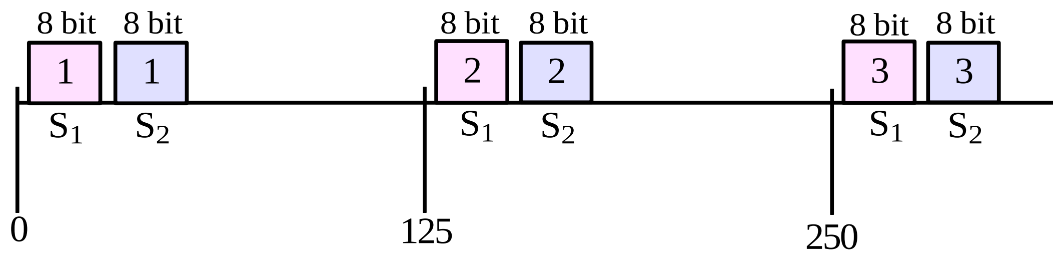
Abbiamo appena detto che, secondo lo standard PCM, il canale usato per la trasmissione deve poter trasmettere almeno **64 kbit/sec**. Nella realtà, i canali utilizzati consentono di ottenere delle capacità di gran lunga maggiori e questo ha suggerito la ricerca di un modo per sfruttare tali grosse prestazioni. La prima cosa che è venuta in mente è il tentativo di sfruttare queste prestazioni per trasmettere *più conversazioni in parallelo*, ossia per trasmettere, utilizzando 1 solo canale fisico, più conversazioni. Vediamo meglio di che si tratta.

Abbiamo detto che la frequenza di campionamento usata nello standard PCM è di 8000(Hz), alla quale corrisponde un periodo di campionamento di 125μsec: questo significa che, una volta prelevato, quantizzato e trasmesso un campione, sono necessari 125μsec perché venga prelevato, quantizzato e trasmesso il campione successivo. Quindi, tra l'invio di un campione e l'invio del successivo, il canale risulta "inattivo" per 125μsec.

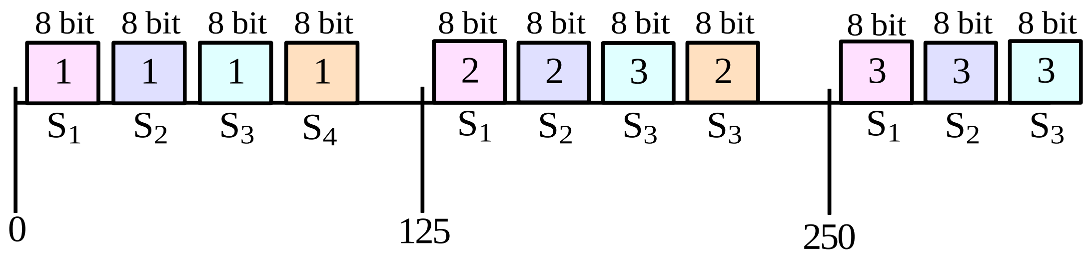
Possiamo visualizzare questa situazione nel modo seguente:



In questo schema, abbiamo rappresentato i tempi di invio del 1° campione, del 2° campione e del 3° campione prelevati dalla generica sorgente S_1. Da questo schema risulta particolarmente evidente il tempo durante il quale il canale risulta inattivo. Allora, possiamo pensare di impiegare il tempo che intercorre tra il primo e il secondo campione della S_1, per inviare il primo campione di un'altra sorgente, che indichiamo con S_2; stesso discorso per l'intervallo di tempo che intercorre tra il secondo ed il terzo campione di S_1 e così via per tutti gli intervalli di tempo:



Se poi avanza ancora tempo, possiamo inviare anche i campioni emessi da una terza sorgente e così via, finche il tempo a disposizione non si esaurisce:



In tal modo quindi, riusciamo a trasmettere, contemporaneamente, più di una conversazione, ossia riusciamo ad implementare, su un unico canale fisico, più di un canale logico.

Questa tecnica prende il nome di **tecnica TDM**, dove questo acronimo sta per **Time Division Multiplexing** (ossia **multiplazione a divisione di tempo**).

E' subito ovvio che il numero di canali logici che possiamo realizzare con un unico canale fisico dipende dalla capacità di tale canale fisico: considerando che ogni canale logico, secondo lo standard PCM, necessita di 64 kbit/sec, avremo bisogno di un canale fisico di almeno 128 kbit/sec per realizzare due canali logici, di un canale fisico di almeno 192 kbit/sec per realizzare 3 canali logici e così via.

A seconda di quanti canali logici noi realizziamo, si parla di un diverso **livello**: per esempio, si parla di **gerarchia del 1° livello** quando i canali logici realizzati sono **32**.

E' bene sottolineare come non tutti i canali logici sono utilizzati per trasmettere comunicazioni telefoniche, in quanto alcuni di essi vengono utilizzati per comunicazioni di servizio, ossia per inviare informazioni necessarie a far funzionare l'apparato di comunicazione (le cosiddette **informazioni di segnalazione**); in particolare, le informazioni riguardano la "sincronizzazione" degli apparati e lo scambio di comunicazioni tra le varie centrali: per esempio, nella gerarchia del 1° livello, il canale numero 1 viene usato per la sincronizzazione, mentre il canale 17 viene usato per mettere in comunicazione le varie centrali; i restanti 30 canali sono dedicati alle conversazioni telefoniche vere e proprie.

Nella seguente tabella sono indicate le caratteristiche fondamentali dei primi 5 livelli utilizzati nella pratica:

LIVELLO	**CAPACITA' DEL CANALE** (Megabit/secondo)	**NUMERO DI CANALI LOGICI**
1	2.048	30+2
2	8.448	120+8
3	34.368	480+32
4	139	1920+128
5	560	7680+256

Facciamo osservare come i normali cavi telefonici appartengono al livello 5, per cui permettono 7680 conversazioni contemporanee.

Codifiche preditive

INTRODUZIONE

Abbiamo detto che il campionamento del segnale telefonico viene effettuato, a livello internazionale, secondo le specifiche dello **standard PCM**, il quale prevede una frequenza di campionamento $f_C = 8000(\text{Hz})$; abbiamo inoltre visto che questa frequenza corrisponde ad un valore del periodo di campionamento $T_C = 125(\mu\,\text{sec})$. si nota dunque che i campioni prelevati dal segnale telefonico sono molto ravvicinati tra loro. Ci si è chiesto allora se esistesse un qualche legame tra due campioni adiacenti o, più in generale, tra il generico campione e alcuni dei campioni precedenti. Effettivamente, <u>è stato trovato che i campioni sono "correlati" tra loro</u> (tra un attimo vedremo quale sia il senso matematico di questa espressione) e, di conseguenza, si è provato a sfruttare questa correlazione al fine di ridurre il numero di bit usati nella quantizzazione con la tecnica TDM (ricordiamo, a questo proposito, che questo numero di bit è 8 e corrisponde appunto al numero di bit associati a ciascun intervallino tra quelli in cui è stato diviso il range dinamico del segnale campionato).

Sottolineiamo, sin da ora, che questo risparmio di bit è poco necessario nelle normali comunicazioni telefoniche via cavo, dove c'è ampia disponibilità di bande di frequenza, mentre è molto più importante nel campo delle **comunicazioni cellulari** (dove invece le frequenze libere diventano sempre di meno ed è inoltre importante trasmettere quanta meno potenza possibile): infatti, sfruttando quello che diremo, sia pure in breve, tra poco, nelle **comunicazioni cellulari** ogni canale logico ha una capacità di circa 10-12 kbit/sec, contro i 64 kbit/sec dei canali tradizionali esaminati in precedenza.

STIMA LINEARE DEI CAMPIONI

Scendendo un po' più nei dettagli, gli studi che sono stati condotti su 2 generici campioni del segnale telefonico hanno condotto al seguente risultato: *la funzione di autocorrelazione tra tali* *campioni risulta avere un valore elevato quando essi sono vicini tra loro, mentre risulta avere valori sempre più bassi man mano che aumenta la loro distanza.*

Questo fatto è stato sfruttato nelle cosiddette **codifiche preditive**: il concetto di fondo è quello per cui, anziché trasmettere, sul canale, il valore di un certo campione, si trasmette un altro valore, molto più piccolo, che corrisponde alla differenza tra il campione stesso ed una sua STIMA, calcolata sulla base dei campioni precedenti (da cui appunto il termine "preditive", nel senso che si cerca di prevedere quale sia il valore del campione conoscendo i valori di alcuni campioni precedenti). Ovviamente, dovendo trasmettere valori numerici molto più piccoli, è ragionevole pensare che ci voglia un numero inferiore di bit, con conseguente velocizzazione del sistema.

Vediamo ad ogni modo, a livello molto qualitativo, come si procede.

Supponiamo che i nostri campioni siano i seguenti:

$$x(1T), x(2T), \dots, x((n-N)T), x((n-N+1)T), \dots\dots, x((n-1)T), x(nT)$$

Per semplicità di notazione, possiamo eliminare il simbolo T, per cui sono

$$x(1), x(2), \dots, x(n-N), x(n-N+1), \dots\dots, x(n-1), x(n)$$

Supponiamo di aver trasmesso i campioni da x(1) a x(n-1) e di dover dunque trasmettere adesso x(n). Allora, prima di effettuare tale trasmissione, calcoliamo una "stima" di x(n), ossia <u>proviamo a determinare il valore di x(n) sulla base dei valori assunti da una parte dei campioni precedenti</u>, per esempio quelli a partire da x(n-N). Evidentemente, tale stima va fatta usando una funzione di tali campioni precedenti. Per semplicità, usiamo una **stima lineare**, ossia valutiamo la stima di x(n), che indichiamo con $\hat{x}(n)$ secondo una semplice combinazione lineare del tipo seguente:

$$\hat{x}(n) = a_1 x(n-1) + a_2 x(n-2) + \dots + a_N x(n-N)$$

Definiamo inoltre la seguente funzione:

$$d(n) = x(n) - \hat{x}(n)$$

Si tratta evidentemente dell'**errore** che noi eventualmente commettiamo tra il valore esatto del campione e la stima da noi effettuata.

Allora, il sistema utilizzato è il seguente: nel momento in cui è disponibile il campione x(n), anziché trasmettere tale campione, si valuta la sua stima $\hat{x}(n)$, si valuta poi d(n) e si trasmette sul canale proprio d(n); in uscita dal canale, viene nuovamente calcolata la stima $\hat{x}(n)$ e quindi, sommando tale stima al valore ricevuto d(n), si ottiene proprio il valore ESATTO di x(n).

Il vantaggio di un meccanismo di questo tipo è evidentemente nel fatto che d(n) ha un valore molto piccolo, tanto più piccolo quanto più è accurata la stima $\hat{x}(n)$, per cui sarà certamente necessario un numero inferiore di bit rispetto a quello necessario per trasmettere direttamente x(n). E' ovvio, inoltre, che sia il trasmettitore sia il ricevitore devono usare lo stesso meccanismo per effettuare le stime, in modo che il valore di $\hat{x}(n)$ sia lo stesso in trasmissione ed in ricezione.

Quindi, la trasmissione procede nel modo seguente:

- vengono prima trasmessi, così come sono, un certo numero di campioni iniziali del segnale;
- a partire da uno di essi (solitamente intorno al decimo), si passa a trasmettere d(n), con conseguente riduzione del numero di bit da inviare e quindi con conseguente aumento della velocità di trasmissione.

E' ovvio, come detto poco fa, che questo metodo ha successo quando d(n) è effettivamente più piccolo di x(n), ossia, in definitiva, quando la stima $\hat{x}(n)$ è molto accurata. L'accuratezza della stima dipende evidentemente dai coefficienti di $\hat{x}(n)$, per cui tutto il problema si riduce a determinare il valore di tali coefficienti tali che venga minimizzato il valore di d(n).

Vediamo allora, a questo proposito, di fare qualche calcolo in più nel caso semplice in cui la stima venga effettuata SOLO sulla base del campione precedente, ossia utilizzando una legge del tipo

$$\boxed{\hat{x}(n) = a_1 x(n-1)}$$

Si parla in questo caso di **stima lineare del primo ordine**, proprio perché si usa solo il campione precedente a quello da stimare.

Il nostro obbiettivo è dunque quello di determinare a_1 in modo da ridurre quanto più è possibile la variabilità di d(n).

Avendo detto che il segnale telefonico può essere considerato e va considerato come un processo stocastico, ciascuno dei campioni da esso prelevati costituirà una variabile aleatoria, per cui anche d(n) sarà una variabile aleatoria; di conseguenza, ridurre la variabilità di d(n) significa ridurre la sua varianza, in quanto sappiamo che la varianza di una variabile aleatoria indica, in pratica, di quanto il valore della variabile si discosta mediamente dal suo valore medio.

Dobbiamo dunque trovare una espressione della varianza di d(n); per fare questo, facciamo due ipotesi semplificative:

- la prima è che il segnale telefonico sia un processo stocastico stazionario, il che significa che esso conserva invariate nel tempo le sue caratteristiche statistiche[3];

- la seconda, senz'altro più realistica della prima, è invece che il segnale telefonico sia un processo a media nulla; questo, in base anche alla stazionarietà, implica che

$$E[x(n)] = E[x(n-1)] = 0$$

e anche che

$$\mathrm{Var}(x(n)) = E\left[x^2(n)\right]$$

Andiamo allora a valutare la media di d(n): dato che $d(n) = x(n) - \hat{x}(n)$ e dato che la media è lineare, possiamo scrivere che

$$E[d(n)] = E[x(n)] - E[\hat{x}(n)]$$

Avendo detto che $E[x(n)] = 0$, quella diventa $E[d(n)] = -E[\hat{x}(n)]$.

Inoltre, essendo $\hat{x}(n) = a_1 x(n-1)$, abbiamo

$$E[d(n)] = -a_1 E[x(n-1)]$$

[3] In precedenza abbiamo ricordato che questa ipotesi non è molto lecita nel caso del segnale telefonico, ma per i nostri ragionamenti ci fa comunque comodo

Ma abbiamo anche detto che $E[x(n-1)]=0$, per cui

$$E[d(n)]=0$$

Abbiamo dunque trovato che, nell'ipotesi per cui il segnale telefonico è stazionario e a media nulla, risulta a media nulla anche d(n). Di conseguenza, la varianza di d(n) sarà semplicemente

$$\mathrm{Var}(d(n)) = E\left[d^2(n)\right] - E^2[d(n)] = E\left[d^2(n)\right]$$

Sostituendo l'espressione di d(n), abbiamo che

$$\mathrm{Var}(d(n)) = E\left[x^2(n) + \hat{x}^2(n) - 2x(n)\hat{x}(n)\right]$$

Sfruttando la linearità della media, abbiamo che

$$\mathrm{Var}(d(n)) = E\left[x^2(n)\right] + E\left[\hat{x}^2(n)\right] - 2E\left[x(n)\hat{x}(n)\right]$$

Sostituendo adesso $\hat{x}(n) = a_1 x(n-1)$, abbiamo che

$$\mathrm{Var}(d(n)) = E\left[x^2(n)\right] + a_1^2 E\left[x^2(n-1)\right] - 2a_1 E\left[x(n)x(n-1)\right]$$

Data l'ipotesi di stazionarietà del segnale telefonico, è chiaro che $E\left[x^2(n)\right] = E\left[x^2(n-1)\right]$, per cui

$$\mathrm{Var}(d(n)) = \left(1+a_1^2\right)E\left[x^2(n)\right] - 2a_1 E\left[x(n)x(n-1)\right]$$

Ora, ci ricordiamo che la **funzione di autocorrelazione** di una generica variabile aleatoria X(t) tempo-continua è quella funzione definita come $R_X(\tau) = E\left[X(t)X(t-\tau)\right]$: se poniamo X=x, t=n (sottointeso T) e τ=1, quella diventa

$$R_X(1) = E\left[x(n)x(n-1)\right]$$

ossia la definizione di funzione di autocorrelazione, calcolata nell'istante 1, di una variabile aleatoria tempo-discreta.

In base a questo, possiamo evidentemente affermare che l'espressione più generale possibile per la varianza di d(n), sotto le ipotesi che stiamo considerando, è

$$\boxed{\mathrm{Var}(d(n)) = \left(1+a_1^2\right)R_X(0) - 2a_1 R_X(1)}$$

Il nostro scopo è quello di trovare il valore di a_1 che minimizza questa quantità: dobbiamo dunque derivare rispetto ad a_1 e imporre che la derivata sia uguale a zero. La derivata è

$$\frac{d}{da_1}\mathrm{Var}(d(n)) = 2a_1 R_X(0) - 2R_X(1)$$

Imponendo che sia nulla, si trova evidentemente che

$$\boxed{a_1^* = \frac{R_X(1)}{R_X(0)}}$$

<u>Questo valore a_1^* del coefficiente è dunque il valore migliore possibile, ossia quello che determina il minimo valore possibile per $\mathrm{Var}(d(n))$</u>. Vediamo allora qual è questo valore minimo: andando a sostituire nella relazione

$$\mathrm{Var}(d(n)) = \left(1 + a_1^2\right) R_X(0) - 2a_1 R_X(1)$$

otteniamo

$$\mathrm{Var}(d(n)) = \left(1 + \frac{R_X^2(1)}{R_X^2(0)}\right) R_X(0) - 2\frac{R_X(1)}{R_X(0)} R_X(1) = \left(R_X(0) + \frac{R_X^2(1)}{R_X(0)}\right) - 2\frac{R_X^2(1)}{R_X(0)} =$$
$$= R_X(0) - \frac{R_X^2(1)}{R_X(0)}$$

Scritto in forma migliore, possiamo affermare che il valore minimo di $\mathrm{Var}(d(n))$ è

$$\boxed{\left[\mathrm{Var}(d(n))\right]_{MIN} = R_X(0)\left(1 - \frac{R_X^2(1)}{R_X^2(0)}\right)}$$

Adesso ci ricordiamo di un’altra definizione data in precedenza, che è quella di **coefficiente di correlazione**: date due variabili aleatorie X ed Y, si definisce *coefficiente di correlazione* tra di esse la quantità

$$\rho_{XY} = \frac{E[XY] - E[X]E[Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Se prendiamo X=x(n) e Y(x(n-1)), ρ diventa

$$\rho(1) = \frac{E[x(n)x(n-1)] - E[x(n)]E[x(n-1)]}{\sqrt{\mathrm{Var}(x(n))\mathrm{Var}(x(n-1))}}$$

Dato però che $E[x(n)] = E[x(n-1)] = 0$, abbiamo anche che

$$\rho(1) = \frac{E[x(n)x(n-1)]}{\sqrt{\mathrm{Var}(x(n))\mathrm{Var}(x(n-1))}}$$

Inoltre, abbiamo detto che $R_X(1) = E[x(n)x(n-1)]$, per cui

$$\rho(1) = \frac{R_X(1)}{\sqrt{\mathrm{Var}(x(n))\mathrm{Var}(x(n-1))}}$$

Inoltre, in base alla stazionarietà, è chiaro che $\mathrm{Var}(x(n)) = \mathrm{Var}(x(n-1)) = E\left[x^2(n)\right] = R_X(0)$, per cui possiamo concludere che il coefficiente di correlazione tra x(n) e x(n-1) vale

$$\boxed{\rho(1) = \frac{R_X(1)}{R_X(0)}}$$

Andando a sostituire nell'espressione di $\left[\mathrm{Var}(d(n))\right]_{MIN}$, possiamo riscrivere quest'ultimo valore nella forma

$$\boxed{\left[\mathrm{Var}(d(n))\right]_{MIN} = R_X(0)\left(1-\rho^2(1)\right)}$$

Esprimere $\left[\mathrm{Var}(d(n))\right]_{MIN}$ in questo modo ci serve in quanto possiamo fare il discorso seguente:

il coefficiente di correlazione tra x(n) ed x(n-1) è un parametro che dipende SOLO dai campioni x(n) e x(n-1), quindi dipende solo dalle caratteristiche del segnale telefonico; dato che esso varia nell'intervallo [-1,+1], è chiaro che il suo quadrato varia nell'intervallo [0,+1]. Di conseguenza, una volta preso per a_1 proprio il valore di ρ(1), noi siamo certi di essere nella situazione migliore possibile, ma <u>ancora dipendiamo dal valore assunto da ρ(1)</u>.

Il caso migliore che ci può capitare è che $\rho^2(1) = 1$, in quanto in questo caso avremmo $\left[\mathrm{Var}(d(n))\right]_{MIN} = 0$. Tuttavia, questo è un caso ideale assolutamente non realizzabile: d'altra parte, dire che la varianza di una variabile aleatoria è nulla significa dire che tale variabile assume sempre lo stesso valore (pari al valore medio), per cui sarebbe del tutto inutile trasmettere d(n).

Al contrario, invece, le cose per noi vanno tanto meglio quanto più $\rho^2(1)$ si approssima ad 1.

GUADAGNO DI PREDIZIONE

Abbiamo detto che la convenienza della tecnica delle "*<u>codifiche preditive</u>*" subentra solo nel caso in cui effettivamente il valore di d(n) risulta inferiore al valore di x(n): solo in questo caso, infatti, noi potremo usare meno bit da inviare sul canale.

Possiamo perfezionare questo concetto introducendo un nuovo parametro, che prende il nome di **guadagno di predizione** e che è definito come rapporto tra la funzione di autocorrelazione di x(n) e la funzione di autocorrelazione di d(n), entrambe calcolate in 0: quindi il guadagno di predizione è

$$\boxed{\frac{R_X(0)}{R_d(0)}}$$

Questo rapporto può essere ulteriormente semplificato sulla base delle ipotesi semplificative fin qui adottate: infatti, noi stiamo supponendo che la stima di x(n) sia lineare del tipo $\hat{x}(n) = a_1 x(n-1)$ e stiamo anche supponendo che il segnale telefonico sia stazionario a media nulla, il che implica che

$$E[x(n)] = E[x(n-1)] = E\left[\hat{x}(n)\right] = E\left[d(n)\right] = 0$$

In base a quest'ultima relazione, si deduce che

$$R_X(0) = E[x^2(n)] = \text{varianza di } x(n)$$
$$R_d(0) = E\left[d^2(n)\right] = \text{varianza di } d(n)$$

per cui il guadagno di predizione diventa in queste ipotesi

$$\frac{R_X(0)}{R_d(0)} = \frac{E\left[x^2(n)\right]}{E\left[d^2(n)\right]}$$

Ovviamente, è possibile parlare rigorosamente di "*guadagno*" solo nel caso in cui quel rapporto è maggiore di 1, in quanto questo significa che la varianza di d(n) è minore di quella di x(n) e quindi noi possiamo effettivamente usare un numero minore di bit durante la trasmissione.

Vediamo ad esempio quanto vale il guadagno di predizione quando prendiamo come valore del coefficiente a_1 proprio il valore migliore possibile, che abbiamo trovato essere

$$a_1^* = \frac{R_X(1)}{R_X(0)} = \rho$$

Abbiamo in precedenza trovato che, in corrispondenza di questo valore del coefficiente a_1, la varianza di d(n) assume il suo valore minimo, che è

$$\left[\mathrm{Var}(d(n))\right]_{MIN} = R_X(0)\left(1-\rho^2(1)\right)$$

(valore minimo che, comunque, dipende ancora dal valore assunto da $\rho(1)$).

Andando allora a sostituire nel rapporto $\frac{R_X(0)}{R_d(0)} = \frac{E\left[x^2(n)\right]}{E\left[d^2(n)\right]}$, abbiamo che

$$\frac{R_X(0)}{R_d(0)} = \frac{R_X(0)}{R_X(0)\left(1-\rho^2(1)\right)} = \frac{1}{1-\rho^2(1)}$$

Dato che il coefficiente di correlazione varia nell'intervallo $[-1,+1]$, è chiaro che il suo quadrato varia nell'intervallo $[0,+1]$, per cui quella quantità può essere maggiore o uguale di 1 a seconda di quale sia il valore di ρ. Il caso peggiore che può capitare è che ρ=1, nel quale caso otteniamo un guadagno di predizione pari proprio ad 1, ossia non otteniamo né un guadagno né una perdita. Questo significa che <u>prendendo il valore a_1^*, male che vada non perdiamo niente, mentre, in genere, otteniamo un guadagno</u>.

Vediamo adesso un altro caso particolare: supponendo sempre che la stima di x(n) viene effettuata secondo la relazione $\hat{x}(n) = a_1 x(n-1)$, vediamo quanto vale il guadagno di predizione nel caso in cui a_1=1, ossia nel caso in cui stimiamo che il nostro campione x(n) abbia lo stesso valore del campione precedente.

Dobbiamo sempre partire dalla definizione, ossia

$$\frac{R_X(0)}{R_d(0)} = \frac{E\left[x^2(n)\right]}{E\left[d^2(n)\right]}$$

Per a_1 generico, abbiamo trovato prima che la varianza di d(n) è

$$Var(d(n)) = E\left[d^2(n)\right] = \left(1+a_1^2\right)R_X(0) - 2a_1R_X(1)$$

per cui

$$\frac{R_X(0)}{R_d(0)} = \frac{R_X(0)}{\left(1+a_1^2\right)R_X(0) - 2a_1R_X(1)}$$

Questa può anche essere riscritta nella forma

$$\frac{R_X(0)}{R_d(0)} = \frac{1}{\left(1+a_1^2\right) - 2a_1\dfrac{R_X(1)}{R_X(0)}} = \frac{1}{1+a_1^2 - 2a_1\rho(1)}$$

Se a_1=1, abbiamo quindi che

$$\left(\frac{R_X(0)}{R_d(0)}\right)_{a_1=1} = \frac{1}{2(1-\rho(1))}$$

Da quest'ultima relazione si deduce che

$$\frac{1}{2(1-\rho(1))} \ge 1 \quad \text{se e solo se } \rho(1) > 0.5$$

Richiedere un valore di $\rho(1)$ maggiore di 0.5 significa richiedere una correlazione particolarmente elevata tra x(n) e x(n-1) ed è dunque l'unica condizione sotto la quale noi possiamo avere un guadagno quando a_1=1.

Schema della tecnica di codifica preditiva

Al fine di riassumere quanto abbiamo fino ad ora detto, possiamo tracciare uno schema sintetico che raffigura i passaggi di cui si compone la tecnica di codifica preditiva:

- in testa a tale schema c'è ovviamente il campione x(n) che viene generato dal campionatore;
- al valore di questo campione dobbiamo sottrarre la sua stima $\hat{x}(n)$: l'esito della sottrazione è il valore d(n) da trasmettere;
- ovviamente, per effettuare la trasmissione, dobbiamo effettuare una quantizzazione di d(n): quindi d(n) entra in ingresso ad un quantizzatore e la sequenza di bit $\tilde{d}(n)$ che ne viene fuori passa nel canale e viene trasmessa;

- una volta arrivata all'apparato ricevente, questa sequenza viene sommata al valore di $\hat{x}(n)$, che viene dunque ricalcolato anche in questa sede, per cui il risultato finale è quello per cui al ricevitore arriva esattamente x(n).

Uno schema di questo tipo prende il nome di **PCM differenziale** (l'acronimo è **DPCM**), proprio perché rispetta tutti gli standard dello schema PCM salvo il fatto di trasmettere d(n) anziché x(n).

Naturalmente, le cose risultano semplificate, anche nel caso in cui la stima sia del tipo

$$\hat{x}(n) = a_1 x(n-1) + a_2 x(n-2) + \ldots + a_N x(n-N)$$

con N>1, quando comunque il segnale telefonico si può ritenere stazionario. Questo, invece, sappiamo che non è generalmente possibile, per cui la determinazione dei coefficienti della stima tali da minimizzare d(n) risulta estremamente complessa. Ciò che si fa è determinare tali coefficienti non in modo univoco, bensì passo per passo, proprio al fine di tenere conto della non-stazionarietà del processo. Si parla in questo caso di **ADPCM**, che significa **PCM Differenziale Adattativo** Questa tecnica è quella attualmente usata, con un elevato grado di sofisticazione, nel campo della telefonia cellulare, ad esempio nel **sistema GSM**.

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: http://users.iol.it/sandry
succursale: http://digilander.iol.it/sandry1