Internet Video Streaming

Manuel Cadeddu

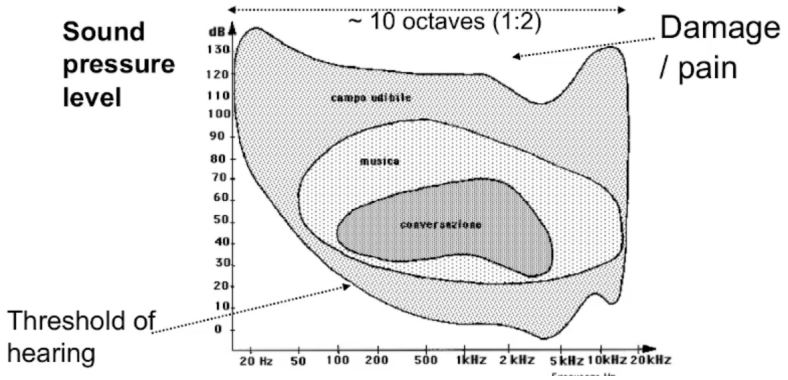
2023-02-28

Indice

1 Segnali Audio e Voce

1.1 Introduzione

Un'**onda acustica** (suono) è una variazione della pressione dell'aria nel tempo. La sua **ampiezza** indentifica il "volume" del suono ed è misurata come differenza tra la pressione locale e quella dell'onda sonora. La sua **frequenza** identifica il fonema, ovvero il suono ("a", "b"). Solo una parte delle onde sonore sono udibili dall'uomo (es. non percepiamo infra/ultra-suoni) e solo una parte di queste è riproducibile tramite le corde vocali (umane).
Come possiamo notare dal seguente grafico, gli umani riescono a sentire suoni con frequenze fino a **20 KHz** e fino a **130 dB**. I suoni riproducibili sono di frequenze comprese tra i **100** e i **5-7 KHz** e tra i **25** e i **70 dB**. Notare che per musica si intende quella prodotta ad esempio da strumenti.



I suoni che superano la soglia superiore del Sound Pressure Level (volume del suono, dato dalla variazione di pressione) causano dolore/danni all'orecchio umano, quelli inferiori alla soglia di udibilità invece non sono udibili.
Dal precedente grafico possiamo fare diverse osservazioni:

- le due scale sono logaritmiche (ogni punto equivale ad un "x val" rispetto al precedente): sono udibili suoni con frequenza maggiore fino a 20k volte rispetto alla frequenza minima e suoni con un valore fino a 10^{13} volte maggiore del minimo udibile;
- $SPL = 10 \log_{10}(P/P_0)$
  - P_0 è il valore minimo percettibile a 1 KHz
- il suono udibile è di circa 10 ottave (raddoppiamento di frequenza);
- quando si lavora in dB si hanno sempre misure relative (non assolute);
- la miglior frequenza per scoltare suoni bassi è a circa 3 KHz.

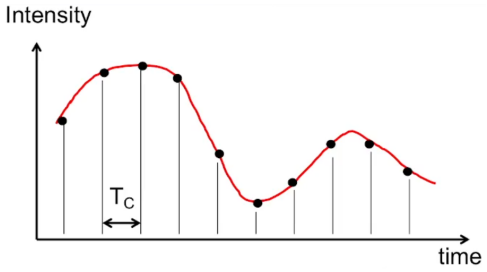
1.2 Converione A/D

La conversione analogico/digitale consiste nella trasformazione di un segnale continuo in uno discreto. La conversione avviene attraverso diverse fasi:

1. cattura del segnale analogico (es. con microfono);
2. **campionamento**;
3. **quantizzazione**.

1.2.1 Campionamento

Il campionamento consiste nella cattura del segnale in precisi istanti di tempo (solitamente ad intervalli regolari). In questo modo si passsa da segnale a tempo continuo a segnale a tempo discreto.

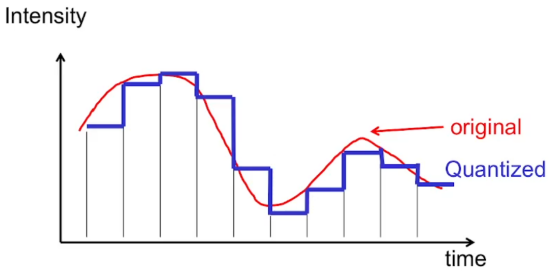


Nota:

- per poter ricostruire il segnale analogico da quello campionato è necessario che la frequenza di campionamento $f_c = 1/T_c$ sia almeno $2 * f_{max}$ del segnale che deve essere campionato (**Teorema di Nyquist**).

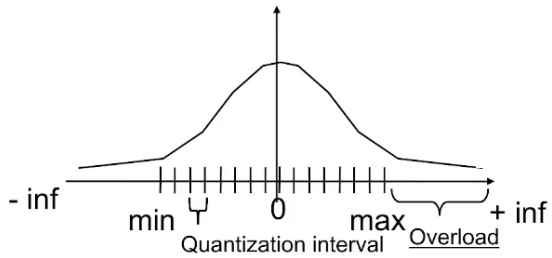
1.2.2 Quantizzazione

La quantizzazione consiste nella mappatura dei valori continui del segnale analogico in valori discreti.



La tecnica più semplice è la **quantizzazione uniforme** (o **lineare**): l'intervallo dei possibili valori viene diviso in intervalli della stessa dimensione.

Notare che spesso viene persa informazione perché si arrotonda il valore del segnale e che, ovunque cada il valore quantizzato, l'accuratezza dell'operazione è sempre la stessa.
Quando si progetta un quantizzatore bisogna fare in modo che il range operativo sia massimo (per catturare "valori estremi") e che la distanza dei livelli di quantizzazione sia minima (per ridurre l'errore di quantizzazione).



In caso di **overload** (il valore del segnale è maggiore del massimo o minore del minimo dell'operational range) il segnale viene associato al valore max o min. Per scegliere la dimensione dell'operational range si usa la seguente regola:

$$\textbf{operational range} = 4\sigma$$

La differenza tra il valore assegnato dal quantizzatore e quello effettivo è detto **errore di quantizzazione**. Considerando questo valore come un **rumore**, è possibile includerlo nel **Rapporto Segnale/Rumore** (**SNR** = **Signal-to-Noise Ratio**, 40-50 dB è già un buon audio). In realtà il SNR è più utile se calcolato comparando la potenza del segnale più che la tensione (in dB):

$$SNR = 10\ log_{10}(\sigma_v^2/\sigma_e^2)$$

- σ_v^2 = varianza del segnale (da gaussiana);
- σ_e^2 = varianza del rumore (dipende dal quantizzatore).

Assumendo che il rumore di quantizzazione abbia distribuzione uniforme, si ha:

$$SNR \sim 6*N - f(sigma_v^2/X_{max}^2)$$

- N = numero bit;
- X_{max} = range di quantizzazione;
- notare che il SNR vale 6*N solo in condizioni ottimali, ovvero solo quando viene sfruttato tutto il range del quantizzatore (quando il quantizzatore ha un range uguale alla varianza del segnale). Per esempio, se si ha un range di 10 V e il segnale in ingresso è di mV, il SNR è alto.

Quando si progetta un quantizzatore bisogna scegliere:

- il range operativo (valori di fondoscale);
- il SNR desiderato quando è usato l'intero range operativo.

1.2.3 PCM (Pulse Code Modulation)

Un **Modulatore a Impulsi Codificati** (PCM) è un dispositivo in grado di campionare il segnale e di quantizzarlo su N bit per campione. Per esempio 12 bit per la telefonia e 16 bit per i CD audio. Maggiore è il numero di bit, maggiore è il SNR e, di conseguenza, il ruomore è meno fastidioso.
Vediamo degli esempi sulla frequenza di trasmissione di un PCM lineare:

- CD Audio: 44100 * 16 * 2 = 1,411,200 bit/s
  - 44100 = 2 * f_{max};
  - 16 = bit quantizzazione;
  - 2 = numero di canali (cassa dx e sx);
- Telefono: 8000 * 12 = 96,000 bit/s;
  - 8000 = 2 * f_{max} (voce udibile, inoltre la frequenza massima è stata leggermente tagliata);
- Video: 720 * 576 * 8 * 30 = 100 Mbit/s;
  - 720 * 576 = risoluzione (pixel);
  - 8 = per colore;
  - 30 = fps;

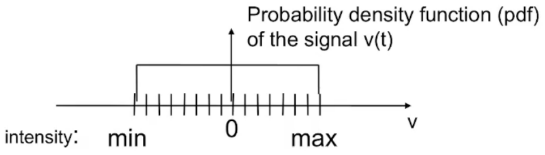
1.3 Codifica della Voce

1.3.1 Voce Telefonica

Come visto in precedenza, la voce umana va dai 100 Hz ai 5-7 kHz circa. Per avere un'applicazione telefonica (intelligibilità = chiarezza interpretativa) decente però, non è necessario coprire interamente questo range, anche se si perde parte della naturalezza del discorso. Per questo vengono trasportate le frequenze tra i **300 Hz** e i **3400 Hz**. I valori inferiori ai 300 Hz vengono tagliati perché più che trasportare informazioni trasportano energia, con conseguenti consumi inutili. I valori oltre i 3400 Hz invece non risulatano necessari per comprendere il messaggio. Inoltre, maggiore è la distanza, minore è la frequenza trasportabile.
Col passaggio dalla telefonia analogica a quella digitale, non sono cambiati questi parametri e si è decisi di usare F_c = **8000** Hz (> 2*3400) e quantizzazione uniforme con valori a **12 bit**. Di conseguenza, la frequenza di trasmissione è di 12*8000 = 96 kbit/s.

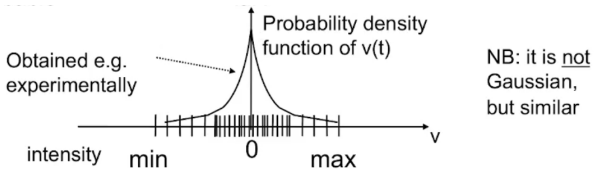
1.3.2 Quantizzatore Uniforme

La tecnica più semplice é la quantizzazione uniforme (o lineare): l'intervallo dei possibili valori viene diviso in intervalli della stessa dimensione. Questa tecnica risulta ottimale (minimizza il rumore di quantizzazione) se il segnale ha densità di probabilità (pdf) uniforme all'interno del range operativo del quantizzatore (le frequenze si presentano tutte con la stessa frequenza), cosa assai rara.

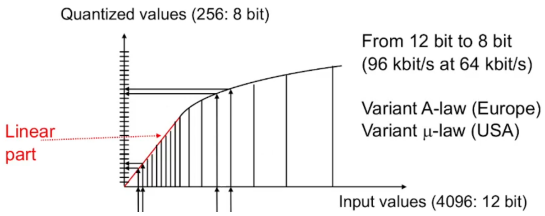


1.3.3 Quantizzatore Non Uniforme

Poichè la pdf della voce umana non è una distribuzione uniforme ma una **distribuzione gamma** (simile a gaussiana ma più "a punta"), è meglio avere più intervalli dove il segnale è più probabile. Per questo motivo è stata introdotta la quantizzazione non uniforme.



Per calcolare l'intervallo della distribuzione che fornisce il massimo SNR viene utilizzato un algoritmo che deriva dal il **teorema di Max-Lloyd**. Applicando questo algoritmo al segnale vocale, si è riusciti ad ottenere lo stesso SNR che si otteneva con 12/16 bit usando solo 8 bit. Per questo, la **ITU** (International Telecommunication Union) ha definito uno standard (**ITU-T G711**): voce telefonica a 64 kbit/s (8 bit, $F_c = 8$ kHz) e quantizzatore **log-PCM** (Pulse Code Modulation). In realtà questo standard non definisce come costruire il quantizzatore ma come passare da quello a 12 bit a quello a 8.
Un codificatore/decodificatore Log-PCM prevede i livelli distribuiti su una scala semi-logaritmica: nella prima parte si ha corrispondenza lineare tra la scala a 12 bit e quella a 8, spostanosi verso gli estremi invece è logaritmica.

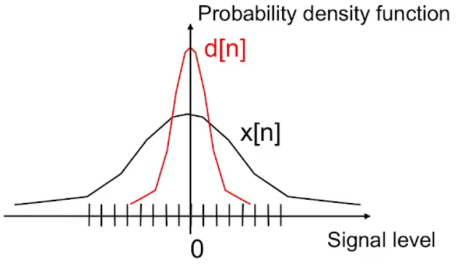


Notare che sono presenti due varianti (una in Europa e una negli USA) e che lo standard definisce come passare da quello uniforme a 12/16 bit e non come costruire il quantizzatore perché quello uniforme è molto più facile da costruire. L'algoritmo per questa conversione risulta molto semplice e utilizza una semplice tabella di lookup per trovare il valore a 8 bit associato a quello di 12/16.
Questo standard è stato utilizzato per la prima volta per le dorsali (molto più semplice gestire segnale digitale) ed è quello ancora usato per quanto rigurda la telefonia fissa (solitamente si trasmette il segnale analogico fino all'armadio di strada e poi viene codificato/decodificato).
I vantaggi di questa tecnologia sono la riduzione dei bit e la semplicità (costo):

- non usa memoria: non usa RAM ma ROM;
- ritardo molto basso: non deve attendere N campioni ma ne codifica uno alla volta;
- statico: non servono algoritmi, processore...

1.3.4 Codifica Differenziale

Per ridurre ulteriormente la banda rispetto al G.711 (64 kbit/s) è possibile ragionare su un gruppo di campioni anziché sul singolo campione. Per esempio, se i campioni sono molto correlati, la differenza tra due campioni consecutivi può essere rappresentata con meno bit rispetto a quelli necessari per trasmettere il valore assoluto dei campioni. La tecnica appena descritta è chiamata "codifica differenziale" e viene implementata nel **Differential PCM** (**DPCM**).
Come possiamo osservare nel seguente grafico, la pdf della differenza tra due segnali ha varianza minore rispetto a quella delle frequenze assolute, di conseguenza è possibile ridurre il numero di bit.



Il lavoro eseguito da codificatore e decodificatore è il seguente:

- **encoder**: d[n] = x[n] - $\hat{x}$[n-1]
- **decoder**: $\hat{x}$[n] = $\hat{d}$[n] + $\hat{x}$[n-1]

Notare che codificatore e decodificatore usano il valore del segnale assegnato dal quantizzatore (accento circonflesso) e non quello del segnale originale.

Predizione nei DPCM

Da ulteriori studi sui segnali, ci si è accorti che non tutti i valori dei campioni $\hat{x}$[n-1] hanno la stessa "importanza", per questo è stato introdotto un **coefficiente di predizione** α.
Le performance del DPCM sono massimizzate quando la varianza dell'errore di predizione è minima:

$$\min_{\alpha} \sigma_d^2 = E[(x[n] - \alpha \cdot x[n-1])^2]$$

Facendo alcuni calcoli (non importanti) si ottiene (importante):

$$d[n] = x[n] - \alpha x[n-1]$$
$$\boldsymbol{min_{\propto}\ \sigma_d^2 = E[d^2[n]]}$$
$$\frac{\partial}{\partial \alpha} E[(x[n] - \alpha x[n-1])^2] = 0$$
$$\frac{\partial}{\partial \alpha} E[x^2[n] - 2\alpha x[n]x[n-1] + \alpha^2 x^2[n-1]] = 0$$
$$E[0 - 2x[n]x[n-1] + 2\alpha x^2[n-1]] = 0$$
$$E[2\alpha x^2[n-1]] = E[2x[n]x[n-1]]$$
$$\Rightarrow \quad \boldsymbol{\alpha = \frac{E[x[n]x[n-1]]}{E[x^2[n-1]]} = \rho}$$

Ovvero, per massimizzare le prestazioni, α deve essere uguale al coefficiente di correlazione ρ.
Notare che α è sempre compreso tra 0 e 1 e che questo è un risultato generale che vale per tutti i tipi di segnali.
Il valore tipico di ρ **per il segnale vocale è 0.9** (il 90% del campione corrente può essere predetto dal precedente).

DPCM di Ordine N-esimo

Il ragionamento del paragrafo precedente può essere esteso considerando N campioni precedenti all'attuale (**DPCM di ordine N-esimo**) per migliorare la predizione:

$$d[n] = x[n] - f(x[n-1], x[n-2], \dots, x[n-N])$$

Se f() è una combinazione lineare:

$$d[n] = x[n] - \sum_{k=1}^{N} (\alpha_k x[n-k])$$

Notare che i coefficienti sono diversi in base al campione.
Per scegliere i valori dei coefficienti si usa l'**algoritmo di Levinson-Durbin**.
Il valore di N tipicamente usato per il segnale vocale in banda telefonica è tra gli 8 e i 12.

1.3.5 Predizione Lineare

I risultati ottenuti fino ad ora possono essere usati per prevedere il prossimo campione senza doverlo campionare e trasmettere la differenza di un campione rispetto alla sua predizione. Questa tecnica è chiamata "Linear Prediction".

$$e[n] = x[n] - \tilde{x}[n]$$

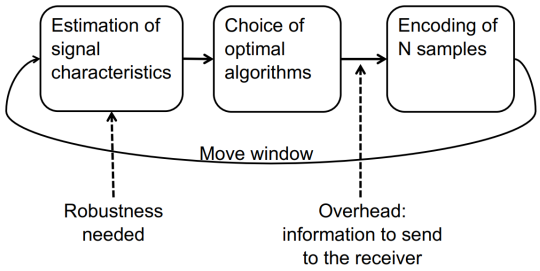
Viene codificato solo l'errore:

La miglior predizione lineare è:

$$\tilde{x}[n] = \sum_{k=1}^{N} (\alpha_k x[n-k])$$

1.3.6 Approccio Statico/Adattivo

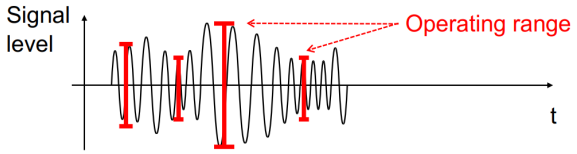
Poiché la correlazione tra campioni può essere diversa a seconda del contesto (es. se uno dice una frase in ambiente ruomoroso e un altro la dice in ambiente silenzioso), ovvero la media dei segnali non è stazionaria, risulta utile poter adattare i parametri
(es. α_k) periodicamente.
Per implementare questa funzionalità il codificatore lavoro come segue:



1. stima delle caratteristiche dei segnali: si esegue analisi per conoscere la correlazione tra i campioni;
2. scelta dell'algoritmo ottimale (che comunque e fisso a seconda del codificatore) e setting dei suoi parametri;
3. codifica dei campioni;
4. il ciclo si ripete con i nuovi campioni.

Notare che la stima delle caratteristiche del segnale deve essere robusta e che l'applicazione di queste tecniche genera overhead.
Le tecniche che mostriamo in questa sezione sono applicabili a tutti i segnali. Considerando il **segnale vocale**, questo può essere considerato **stazionario** (le caratteristiche statistiche non cambiano) **in un intervallo di 5-20 ms** (durata di un fonema).
Quando si progettano tecniche PCM adattive bisogna gestire la *frequenza di adattamento* (50-200 volte al secondo) e l'*overhead* generato (rappresentare i parametri del sistema con meno bit possibili). Per la voce telefonica, un buon compromesso è quindi quello di scegliere $\mathbf{N = 160}$ (20 ms).

Energy-Adaptive PCM Poiché ci sono momenti in cui il sgnale ha volume più o meno alto, si può adattare il range del quantizzatore in base ad esso, riducendo il numero di bit di overhead.



Algoritmi APCM (Adaptive PCM) Gli algoritmi APCM possono adottare due approcci:

- mandare le informazioni sull'adattamento esplicitamente: **feed-forward APCM** (adattamento esplicito);
- non inviare le informazioni sui parametri esplicitamente: **feed-back APCM** (implicito). In questo caso codificatore e decodificatore li calcolano autonomamente. I vantaggi di questa tecnica stanno nel fatto che vengono trasmesse meno informazioni e che analizzando i campioni nel passato si ha più precisione, lo svantaggio che non possono essere fatte predizioni (velocità) ma i codificatore deve aspettare un insieme di campioni prima della decodifica.

APCM e DPCM La tecnica DPCM e quella APCM possono essere utilizzate contemporanemente per predirre coefficienti localmente ottimi. Questo è l'approccio usaro dal codificatore/decodificatore **ADPCM** (**ITU-T G.726 ADPCM**):

- **32 kbit/s** (prima erano 64 kbit/s);
- è aumentata la complessità per calcolare α_k(1 MIPS, Milion Instuctions Per Second): non c'è più solo una tabella di lookup ma servono memoria, processore...

- l'overhead (per la trasmissione degli α_k) è di circa l'8%;

  20ms = 50 times/sec; 10 α_k = 500 numbers = 2000-2500 bits/sec (out of 32000 = 8%)

- anche se sono state sviluppate soluzioni migliori (es. per telefonia mobile), questa tecnica è quella usata per la telefonia fissa;
- l'ADPCM è il massimo raggiungibilie utilizzando le tecniche PCM.