

Codificació de font

Margarita Cabrera
Francesc Tarrés

PID_00185013

Índex

Introducció	5
Objectius	7
1. Fonts de missatges i codificació de font	9
2. Codificació de font i codificació de canal	11
3. Codificació PCM de fonts analògiques	13
3.1. Exemple 1. Codificació PCM de senyals d'àudio: format CD-àudio	16
3.2. Exemple 2. Codificació PCM d'imatges: format <i>bitmap</i>	17
3.3. Exemple 3. Codificació PCM de vídeo: format ITU-601, Norma 4:2:0	19
4. Compresió de fonts. Necessitat i conceptes bàsics	21
4.1. La transformació de les dades	21
4.2. Quantificador	22
4.3. Codificador entròpic	24
5. Mesura de la informació	27
5.1. Informació associada a un missatge	27
5.2. Entropia d'una font de missatges	30
6. Codis de longitud variable. Conceptes bàsics	32
7. Codis de Huffman	34
7.1. Obtenció sistemàtica dels codis de Huffman	34
7.2. Codis de Huffman truncats	38
8. Transformacions de dades	40
8.1. Codificació diferencial. Aplicació en la codificació d'imatge	40
8.2. Predicció lineal. Una generalització de la codificació diferencial ...	43
8.3. Codificació diferencial i predicció d'un senyal d'àudio	44
8.4. Codificació diferencial adaptativa	49
9. Codificació de veu en telefonia mòbil GSM	50
Activitats	53
Bibliografia	54

Introducció

L'objectiu dels sistemes de comunicacions digitals és enviar, amb la màxima **fiabilitat** i **eficiència**, una seqüència de bits que proporcionen un flux d'informació des de l'emissor al receptor. La major part dels continguts d'aquesta assignatura se centren en l'estudi de sistemes i procediments que permeten millorar aquesta fiabilitat i eficiència del sistema de comunicació, entenent que volem transmetre un màxim de bits per segon (eficiència) amb la mínima degradació (fiabilitat). Per tant, gairebé sempre suposarem que el nostre objectiu és transmetre dades sense qüestionar-nos l'origen d'aquestes ni les característiques de la font que representen.

Per a optimitzar el sistema de comunicacions complet és convenient utilitzar procediments que permetin realitzar una representació **eficient** de la informació en forma de seqüències de bits. En alguns casos, la mateixa informació que volem transmetre és de naturalesa digital (fitxers de dades, documents de text, programes informàtics, etc.), per la qual cosa l'obtenció de la seqüència digital és directa. En canvi, en altres aplicacions és necessari definir un procés que ens permeti representar una informació de naturalesa contínua mitjançant una seqüència de bits. A més, volem que aquesta representació sigui eficient, per la qual cosa haurem d'intentar **compactar (comprimir)** la seqüència de bits que representa la informació. Així, doncs, entenem com una representació eficient del missatge la reducció del nombre de bits que s'han d'usar per a la transmissió d'aquest missatge.

Aquesta última afirmació pot tenir alguns matisos en funció de les aplicacions. En efecte, en alguns casos és possible que la càrrega computacional necessària perquè l'equip transmissor pugui realitzar la compressió dels missatges sigui excessiva, la qual cosa augmentaria el cost del transmissor més enllà de la seva possible explotació comercial. En altres aplicacions és possible que hi hagi un canal amb la capacitat suficient per a transmetre la informació sense necessitat de comprimir-la. En tot cas, encara que el grau de compressió final que s'aplica als missatges depèn d'un gran nombre de factors, és inqüestionable que hem de conèixer les tècniques que permeten realitzar aquesta compressió amb la major efectivitat possible.

En aquest mòdul, estudiarem el procés mitjançant el qual una informació de naturalesa analògica es converteix en una seqüència de bits, i posteriorment analitzarem diverses tècniques per a representar de manera compacta aquesta seqüència de bits. Per verificar que la nostra representació de la informació és l'adequada, haurem d'establir mecanismes per a "mesurar" la quantitat d'informació proporcionada per un missatge. Amb aquest objectiu, introduïm alguns conceptes elementals de la teoria de la informació com la definició

formal d'*informació*, l'*entropia* (o informació mitjana proporcionada per una font) i la representació dels missatges mitjançant codis de longitud variable. Finalment, s'introdueixen alguns exemples senzills de sistemes de compressió d'informació per a senyals d'àudio i vídeo.

Objectius

Els objectius que ha d'assolir l'estudiant amb aquest mòdul didàctic són els següents:

1. Comprendre el concepte de codificació font de forma eficient.
2. Saber situar la codificació del missatge dins d'un sistema de comunicacions genèric.
3. Conèixer els elements i blocs integrants que intervenen en un compressor.
4. Comprendre els fonaments de la representació d'informació mitjançant seqüències de bits: PCM.
5. Distingir entre entropia i taxa de bits mitjana d'un codi.
6. Saber calcular codis de longitud variable.
7. Comprendre el procés de codificació i descodificació mitjançant codis de longitud variable.
8. Entendre els mecanismes bàsics per a la compressió d'informació en sistemes diferencials i predictius.
9. Conèixer els principis de codificació del senyal de veu en sistemes de telefonia GSM.

1. Fonts de missatges i codificació de font

En aquest apartat, veurem que la naturalesa dels missatges que volem transmetre pot ser analògica o digital. Els sistemes de comunicacions digitals estan pensats per a transmetre informació en format digital, per la qual cosa, quan la font es presenta en aquest format, es pot transmetre directament. No obstant això, quan la font se'ns presenti en format analògic, necessitarem definir un procediment per convertir-la en digital. En aquest apartat, es defineixen les fonts analògiques i els elements que intervenen en la conversió i codificació d'aquestes a formats digitals.

Les fonts que generen la informació poden ser de naturalesa analògica o digital. Les **fonts digitals** poden procedir de documents de text, fulls de càlcul, programes o qualsevol altre tipus de dades informàtiques. La seva característica principal és que es representen, de manera natural, com a seqüències de bits i que, per tant, es poden transmetre directament mitjançant un sistema de comunicacions digital.

En el cas de les **fonts analògiques**, el missatge que es vol transmetre sol ser una variable real que varia de manera contínua en el temps (o una altra variable independent). Els exemples més típics serien un senyal de veu o d'àudio, la tensió proporcionada per un sensor de temperatura, el nivell de gris o d'un component de color en una imatge, etc.

A la figura 1, es representa un senyal analògic típic que es correspon amb el nivell de tensió obtingut a la sortida d'un micròfon durant un interval de temps de 120 ms. La característica principal dels senyals analògics és que són funcions reals de variable real* que representen un fenomen o magnitud física. Per al nostre exemple, això significa que, en qualsevol instant de temps, el sistema de captura del senyal de veu proporciona un nivell de tensió que pot prendre qualsevol valor real. A partir d'aquesta definició, és evident que els senyals analògics s'hauran de convertir en seqüències de bits abans de poder ser transmeses mitjançant un sistema de comunicacions digital.

* Aquesta definició es pot estendre al cas complex, però no ho considerarem en aquesta assignatura.

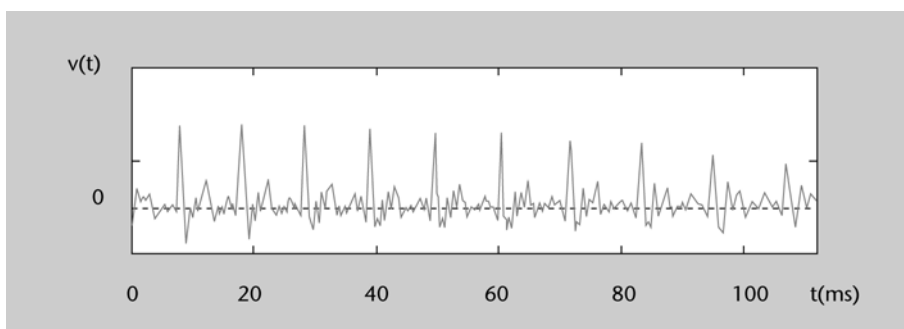


Figura 1. Representació d'un senyal analògic corresponent a un fragment de 120 ms d'un senyal de veu (vocal /o/)

Un **senyal analògic** es pot considerar com una funció real de variable real que representa un fenomen o magnitud física.

2. Codificació de font i codificació de canal

A la figura 2, es representa un diagrama de blocs conceptual amb els elements bàsics que intervenen en un sistema de comunicacions digital. La primera etapa de l'emissor consisteix a convertir qualsevol font de senyal, analògica o digital, en una seqüència de díigits binaris aptes per a la transmissió. Aquesta etapa es coneix com a **codificació de font** o també com a **compressió de dades**, i consisteix a extreure tota la redundància possible del senyal reduint el nombre de bits amb què es representa la informació, sense que això impliqui una pèrdua aparent de qualitat per al receptor final del missatge.

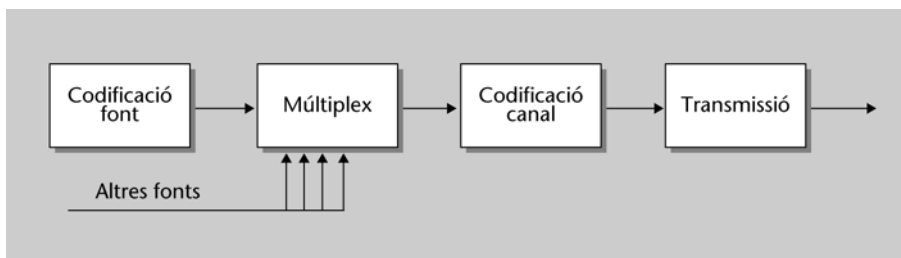


Figura 2. Elements bàsics d'un sistema de comunicacions digitals

El procés de codificació d'una font analògica és més complex que el d'una font digital. En aquest últim cas, si es considera necessari, la codificació de font consisteix a comprimir el flux de bits de la font digital original.

En el cas de fonts analògiques, és necessari definir primer un procediment per a convertir la informació analògica en digital, la qual cosa sempre comporta una possible pèrdua de qualitat o, en altres paraules, la introducció de **soroll** o **distorsió**. En l'apartat següent, descriurem breument la codificació PCM, que és el procediment més utilitzat per convertir una font analògica en digital.

Una vegada el senyal analògic s'ha convertit en digital, encara és possible processar-lo per a reduir el nombre de bits que el representen. Els mètodes i estratègies emprats per a la compressió del senyal són molt variats i és difícil establir un procediment òptim, ja que els resultats depenen, en gran mesura, de les característiques dels senyals. Generalment, els mètodes de compressió avançats combinen simultàniament diferents mètodes per a codificar el senyal.

La **codificació de font** és el procés pel qual una font digital o analògica es converteix en una seqüència de díigits binaris que la representa. Un dels objectius de la codificació de font és extreure la redundància de la font per representar-la de la manera més eficient possible.

El bloc de **multiplexació de font** inclòs a la figura 2 només és present en alguns sistemes de comunicació digital que, generalment, són serveis audiovisuals avançats. El concepte de multiplexació és simple: les trames de bits procedents de fonts diferents (àudio, vídeo) comparteixen un mateix recurs del mitjà de transmissió.

Un exemple típic en el qual es realitza la multiplexació de diverses fonts és la televisió digital terrestre (TDT). En TDT, un canal proporciona informació sobre diversos programes de televisió, mentre que en televisió analògica cada programa està associat a un únic canal. Així, per exemple, la informació que es transmet per un canal de TDT inclou diverses fonts de missatges: quatre o cinc fonts de vídeo (una per programa), diverses fonts d'àudio multicanal amb la banda sonora i la versió original de cada programa, fonts d'àudio amb les versions doblades, informació de dades de teletext avançat, dades interactives, etc. El conjunt de totes aquestes fonts multiplexades s'anomena **flux de transport** (TS o *transport stream*).

És important distingir entre la multiplexació de fonts i la multiplexació de canals o de portadores. La TDT pot ser un bon exemple per a facilitar aquesta distinció. En efecte, les portadores estan multiplexades en l'espectre radioelèctric i es corresponen amb els programes de TV analògica convencional. Per canviar el programa en TV analògica, s'ha de sintonitzar en el receptor una portadora (canal) diferent. En canvi, en TDT cada portadora conté diferents programes de televisió amb les fonts multiplexades.

Tornant a la figura 2, trobem el bloc de la **codificació de canal**. Aquesta codificació s'utilitza per a condicionar els senyals a les característiques del canal. En aquesta etapa s'introdueixen codis de protecció d'errors, codis d'enciptació de la informació, i s'adequa la forma d'ona dels senyals perquè el bloc de modulació o transmissió els pugui utilitzar. Es tracta, per tant, d'adaptar les dades a les característiques del mitjà de transmissió per garantir una correcta recepció dels senyals. El mitjà de transmissió condiona el tipus de **modulació** que s'utilitza i la codificació de canal. Així, en un mitjà com el cable, els sistemes de detecció i correcció d'errors poden ser més simples que en el cas de la comunicació per satèl·lit perquè la relació senyal-soroll és més gran i, per tant, es redueix la probabilitat d'error.

Exemple

La portadora TDT de Televisió Espanyola conté els programes de TVE1, TVE2, 24 H TVE, Clan 50 i Teledporte.

3. Codificació PCM de fonts analògiques

En aquest apartat, veurem el procés més utilitzat per a convertir un senyal analògic en digital, conegut com a *codificació PCM*. L'esquema bàsic d'aquest procediment es mostra a la figura 3.

El senyal analògic original s'aproxima mitjançant una retícula rectangular. La separació entre els elements de la retícula en l'eix d'abscisses és constant i es coneix com el **període de mostreig** del senyal. El teorema del mostreig estableix que, si el nombre de mostres per segon és major que el doble de l'amplada de banda del senyal, aquesta es podrà recuperar a partir de les seves mostres sense ambigüitats. Amb aquestes condicions, en principi, la representació del senyal no millora encara que augmentem la densitat de mostres en l'eix temporal.

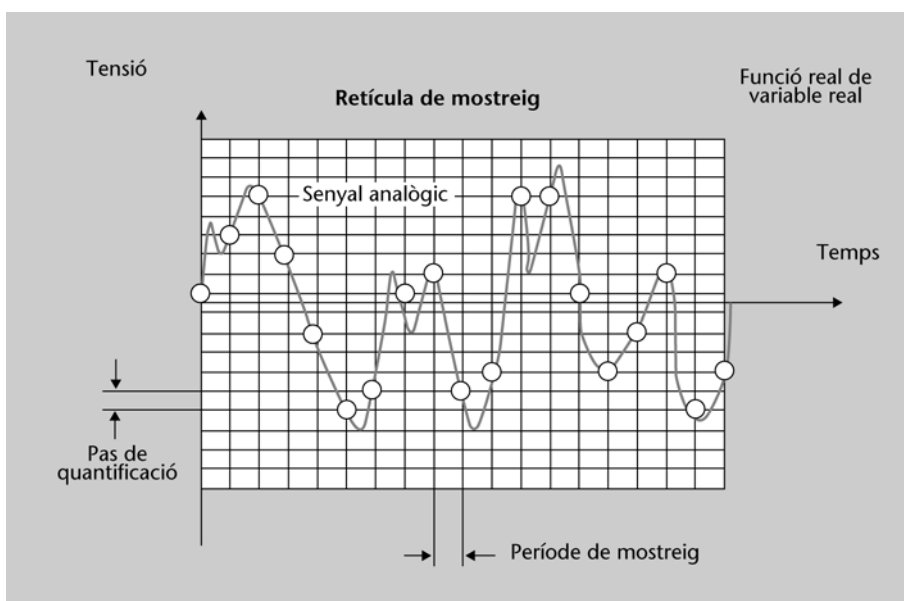


Figura 3. Mostreig i quantificació d'una font analògica

L'interval de separació en l'eix vertical s'anomena **pas de quantificació**. En aquest cas, en reduir l'interval entre dos passos de quantificació consecutius, es millora la representació del senyal. Un interval de quantificació excessivament elevat introdueix un error significatiu en la representació del senyal que, a efectes pràctics, es tradueix en la presència d'efectes indesitjats. Aquests efectes es coneixen com a **soroll de quantificació** i es mostren gràficament a la figura 4. El nom de *soroll de quantificació* prové d'aproximar el nivell del senyal per un dels possibles nivells de quantificació, la qual cosa es pot considerar equivalent a haver sumat un soroll al senyal.

Una vegada fixat el pas de quantificació i els nivells màxim i mínim que pot prendre el senyal, també queda determinat el nombre total de possibles nivells

de quantificació que s'utilitzaran per a la codificació. Ara bé, com el nombre de nivells de quantificació és finit, podem associar una paraula codi diferent a cada un (alfabet finit) i, per tant, el valor del senyal es pot especificar mitjançant paraules codi d'un alfabet finit.

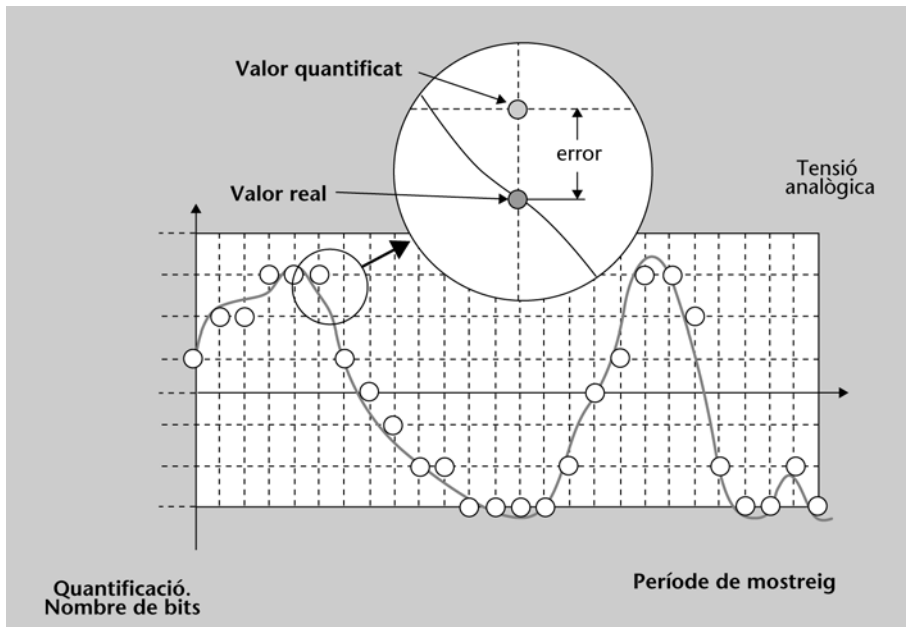


Figura 4. L'error que es comet en aproximar el valor del senyal per un dels nivells de quantificació és equivalent a introduir un soroll (soroll de quantificació).

La figura 5 mostra aquest procés de manera simplificada utilitzant un alfabet reduït de vuit possibles nivells de quantificació. Cada nivell de quantificació té assignat un codi binari (paraula codi). La transmissió d'aquest codi binari proporciona informació del nivell de tensió del senyal en cada instant de mostreig. Aquest procés de codificació de la font es coneix com a **PCM** (*pulse, code modulation*).

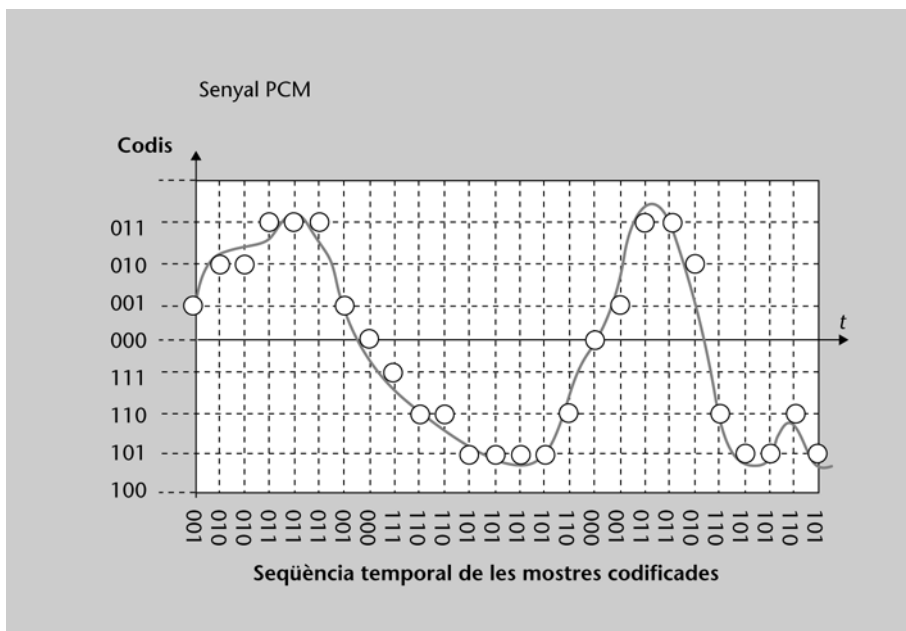


Figura 5. Codificació d'una font analògica en PCM

El nombre de nivells del quantificador (L) es tria com una potència de 2. D'aquesta manera, cada nivell de quantificació es pot associar a una paraula codi de B bits. La relació entre el nombre de nivells del quantificador (L) i el nombre de bits per mostra (B) queda representada per la igualtat següent:

$$L = 2^B .$$

Exemple

En un sistema PCM de 16 bits per mostra, el senyal analògic es quantifica en $2^{16} = 65.536$ nivells diferents.

Mitjançant aquest procés de mostratge i quantificació, la font analògica original s'ha convertit en una font digital. En efecte, en realitzar el mostratge temporal, el senyal analògic es converteix en una seqüència de mostres. El valor d'amplitud de cada mostra es pot codificar mitjançant la selecció d'una paraula codi entre un alfabet finit, de manera que la transmissió del missatge es converteix en la transmissió d'una seqüència determinada de paraules codi. En l'exemple de la figura 5, podríem transmetre una aproximació a la font analògica original enviant la seqüència següent de missatges 001 010 010 011 011... (que són els valors que prenen els codis del senyal en l'eix horitzontal de la figura).

En la pràctica, el pas de quantificació idoni depèn de les característiques del receptor del missatge, que per a senyals d'àudio i vídeo sol ser l'ésser humà. Considerem, com a exemple, la digitalització d'un senyal d'àudio per a aplicacions de reproducció de so en alta fidelitat. En aquest cas, les característiques del sistema auditiu humà són les que determinen el pas de quantificació màxim perquè el senyal es pugui reproduir sense que s'aprecii cap diferència en relació amb el senyal original. El valor del pas de quantificació òptim es determina realitzant proves exhaustives de qualitat subjectiva obtingudes amb diferents oients i intervals de quantificació. El principi bàsic és que no té sentit reduir el pas de quantificació més enllà de les capacitats del sistema auditiu.

En cas que el destinatari final de la informació no sigui l'ésser humà, el procediment és semblant. Considerem com a exemple un sistema de visió artificial que ha de realitzar el reconeixement òptic de matrícules d'automòbil. Ara, el criteri per a determinar el pas de quantificació és que les capacitats de reconeixement del sistema no quedin limitades pel fet d'utilitzar un nombre de nivells excessivament reduït. L'interval de quantificació adequat és, novament, el valor màxim que permet mantenir les prestacions de reconeixement de caràcters del sistema.

En els casos considerats, les prestacions del sistema no queden afectades si s'utilitza un pas de quantificació menor que el que hem definit com a idoni. No obstant això, reduir el pas de quantificació implica augmentar el nombre de bits, la qual cosa en general representa un increment en el cost tecnològic del sistema. En efecte, a continuació veurem alguns exemples de codificació en PCM de senyals d'àudio, imatges o vídeo, i per a cada un determinarem el nombre de bits associat a la font. Veurem que, en tots els casos, la codificació

de la font en PCM implica un nombre de bits molt elevat que serà convenient comprimir si volem transmetre la informació de manera eficient.

3.1. Exemple 1. Codificació PCM de senyals d'àudio: format CD-àudio

El format CD-àudio (*compact disc digital audio*) va ser el primer sistema que va popularitzar el consum d'informació audiovisual en format digital. L'aparició de l'estàndard és el resultat d'una estreta col·laboració entre les empreses Philips i Sony, en la qual es van definir les característiques tècniques del sistema. Els objectius inicials eren aconseguir un sistema de reproducció d'àudio d'alta qualitat i que pogués suportar un gran nombre de reproduccions sense deteriorar-se, ja que aquest era el principal problema dels discs de vinil. En pocs anys, el sistema CD-àudio es va convertir en l'estàndard de la indústria discogràfica i va desplaçar gairebé per complet els discs de vinil i les *compact cassette*.

El sistema està basat en una digitalització directa del senyal d'àudio analògic mitjançant una codificació PCM. Així, doncs, s'havia d'especificar a quina freqüència es prendrien les mostres del senyal d'àudio i amb quants nivells de quantificació es codificaria cada mostra. Per a la freqüència de mostratge, es van fixar 44.100 Hz (mostres per segon). Aquest nombre compleix amb el **criteri de Nyquist**, ja que és superior al doble de l'amplada de banda del senyal d'àudio, que es pot considerar de 20 kHz. El valor concret de 44.100 mostres per segon es deu que els primers sistemes de registre d'àudio digital, creats per la BBC al final de la dècada de 1960, ja van utilitzar aquesta freqüència de mostratge.

Per a determinar el nombre de nivells de quantificació, es van realitzar nombroses proves de qualitat amb especialistes en àudio. Es va acceptar que, amb un total de 14 bits (16.384 nivells = 2^{14}), es podia obtenir una qualitat d'àudio indistingible dels originals analògics, encara que finalment es va acordar prendre un total de 16 bits per mostra com a estàndard ($2^{16} = 65.536$ nivells). Les paraules de 16 bits (2 bytes), a més d'oferir una qualitat teòrica millor, són compatibles amb totes les memòries i processadors digitals. No obstant això, en la primera època de comercialització de discs compactes, alguns enregistraments es van realitzar amb 14 bits a causa de la complexitat tecnològica associada a les precisions de 16 bits (els dos bits de menor pes eren sempre a zero).

La **taxa de bits** (*bitrate*) d'un senyal PCM es defineix com el nombre mitjà de bits que es requereixen per a codificar un fragment amb una durada d'un segon. La taxa de bits es determina com el producte entre la freqüència de mostratge i el nombre de bits per mostra:

$$r_b = f_m \cdot B.$$

Tenint en compte la freqüència de mostratge, el nombre de bits per mostra, i que el sistema CD-àudio utilitza dos canals d'informació (estèreo), podem calcular la taxa de bits que implica l'emmagatzematge o la transmissió d'aquesta informació:

$$r_b = 2 \text{ canals} \times \frac{44.100 \text{ muestras}}{\text{canal}} \times \frac{16 \text{ bits}}{\text{muestra}} = 1.411.200 \text{ bps} .$$

El valor d'1,4 Mbps obtingut fa que el sistema de CD-àudio no es pugui utilitzar de manera pràctica en un gran nombre d'aplicacions en les quals l'amplada de banda del canal de transmissió o la capacitat de memòria del sistema siguin recursos escassos. Per això, han aparegut un gran nombre d'alternatives per intentar comprimir el flux total de bits sense que s'apreciï una disminució de la qualitat del senyal. Entre aquestes alternatives, destaquen sistemes com el disquet, MP3, AAC (*advanced audio coding*), etc. Amb el sistema de compressió MP3 o l'AAC, podem aconseguir la mateixa qualitat perceptiva que amb CD-àudio amb taxes de 128.000 bps per a senyals estereofònics, que significa una compressió aproximada en un factor 12 (el senyal comprimit ocupa una dotzena part del senyal original).

3.2. Exemple 2. Codificació PCM d'imatges: format *bitmap*

La digitalització d'una imatge també consisteix en un mostratge i una quantificació. En aquest cas, el mostratge es realitza en el pla d'imatge i queda determinat pel nombre de files i de columnes en les quals es descompon una imatge. La imatge es pot interpretar com una matriu, tal com es representa a la figura 6. Els elements de la matriu s'anomenen **elements d'imatge** o **píxels** (*picture elements*).

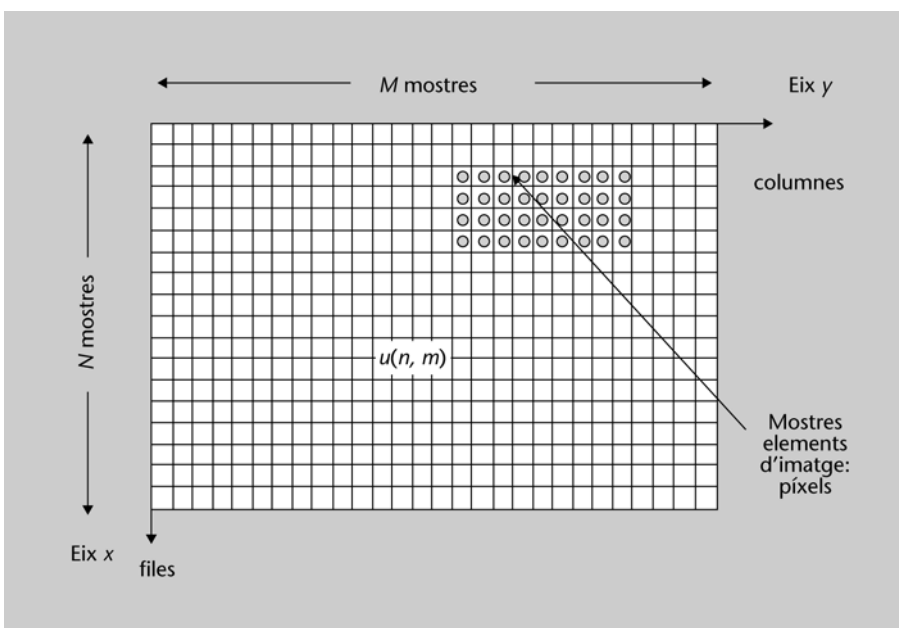


Figura 6. Interpretació d'una imatge com una matriu

El nombre de files i columnes en les quals es descompon una imatge depèn de l'aplicació. En una televisió de definició convencional, es prenen 576 files × 720 columnes (Europa); en una televisió d'alta definició, 1.080 files × 1.920 columnes; mentre que en una càmera de fotografia de 6 Mpíxels prendrien unes 2.000 files × 3.000 columnes.

De la mateixa manera, el nombre de nivells en els quals es quantifica cada element d'imatge (píxel) depèn de les característiques de l'aplicació. En blanc i negre, el sistema visual només pot distingir, aproximadament, entre quaranta i quaranta-cinc nivells de gris diferents. En la pràctica, les imatges en blanc i negre es quantifiquen amb 256 nivells de gris (8 bits) que, encara que són més que els necessaris, facilita l'estructuració de la memòria i els fitxers associats a imatges en paraules de 8 bits. Aquest format de codificació PCM es coneix com a *mapa de bits* o **bitmap de 8 bits**, i s'utilitza per a imatges en blanc i negre. La codificació d'imatges en color se sol realitzar prenent també 256 nivells per a cada un dels components primaris: vermell, verd i blau (RGB, *red, green, blue*). Aquesta codificació dóna lloc a les anomenades *imatges de 24 bits per píxel* o **bitmap de 24 bits** (8 bits per a cada component de color). Hi ha altres formats alternatius com el *bitmap* de 15 bits o el de 16 bits, encara que cada vegada són menys utilitzats. Així, per exemple, el format *bitmap* de 16 bits consisteix a codificar els components de vermell i blau amb 5 bits, mentre que el component verd està codificat amb 6 bits. Els **bitmaps de 32 bits** codifiquen cada component amb 8 bits, i utilitzen 8 bits addicionals per a codificar la "transparència" de la imatge (grau en què la imatge se superposa amb el fons).

El nombre de bits necessari per a emmagatzemar una imatge sense comprimir es pot calcular com el nombre total de píxels, és a dir, el nombre de files pel nombre de columnes ($N_{files} \times M_{col}$) pel nombre de bits que ocupa cada píxel (bpp o bits per píxel)

$$N_{bits} = N_{files} \cdot M_{col} \cdot bpp$$

Com a exemple per a determinar el nombre total de bits que pot ocupar una fotografia, calcularem la mida de memòria necessària per a emmagatzemar o transmetre una imatge en color presa per una càmera de fotografia digital de 6 Mpíxels.

$$N_{imatge} = 2.000 \text{ files} \times 3.000 \text{ columnes} \times 24 \text{ bits} = 144.000.000 \text{ bits} .$$

Novament, el nombre de bits resultant és considerable (trigaríem aproximadament 2,5 minuts a transmetre aquesta imatge per una línia ADSL d'1 Mbps), la qual cosa ens torna a indicar la importància de comprimir la informació abans d'emmagatzemar-la o transmetre-la. Els formats de compressió d'imatge més populars són el JPEG i el TIFF comprimit. El JPEG permet comprimir amb

factors de l'ordre de 10 o superiors sense que es puguin apreciar diferències amb els arxius originals en PCM.

3.3. Exemple 3. Codificació PCM de vídeo: format ITU-601, Norma 4:2:0

Considerarem ara la digitalització d'un senyal de vídeo i calcularem el nombre total de bits que requereix l'emmagatzematge o transmissió d'una seqüència de vídeo digital sense comprimir.

L'estàndard ITU-601 defineix diferents maneres per a codificar el senyal de vídeo digital sense comprimir. Aquestes maneres depenen de l'estàndard analògic de procedència (NTSC, PAL) i de les aplicacions a les quals s'orienta (estudis de televisió, difusió audiovisual). Les normes més utilitzades actualment són la 4:2:2 i la 4:2:0. La primera s'utilitza principalment en estudis de televisió i per a l'intercanvi de materials entre productores. La norma 4:2:0 és la que s'utilitza en la difusió de senyals de televisió o DVD.

Els sistemes de televisió en color, per qüestions de compatibilitat amb els sistemes en blanc i negre, no transmeten els tres components primaris (R, G i B) sinó el component de luminància (i senyal de blanc i negre) més dos components de cromà addicionals (Cr i Cb), que proporcionen la informació de color al receptor (components de diferència de color). El receptor de color pot calcular els components RGB a partir dels senyals YCrCb rebuts. Fixeu-vos que els components RGB i els components YCrCb es poden considerar com dos sistemes de coordenades de color diferents, i cada un té tres variables independents. Podrem passar d'un sistema de coordenades a l'altre mitjançant matrius de canvis de base (sistemes d'equacions).

El sistema de coordenades basat en els components YCrCb és més útil que el sistema RGB principalment per dos motius:

- 1) Per qüestions de compatibilitat amb els antics sistemes de televisió en blanc i negre, que ens obligaven a transmetre la informació de luminància.
- 2) El més important, el sistema visual humà en presenta una major resolució al component de luminància que als components de color.

Podem comprovar aquest últim fet situant dos fils de colors diferents a certa distància d'un espectador. Si la distància és l'adequada, comprovarem que l'espectador és capaç de discriminar que hi ha dos fils, però no en pot distingir els colors. Aquest experiment ens permet inferir que haurem "tractar" millor el component de luminància (és a dir, dedicar-li més bits) que als components de cromà.

Les normes 4:2:2 i 4:2:0 que hem comentat anteriorment utilitzen aquest principi per a emmagatzemar/transmetre una quantitat d'informació menor per als components de cromina que per als de luminància. En concret, la norma 4:2:2 dedica la meitat de la informació a cada un dels components de cromina respecte a la luminància. La norma 4:2:0 dedica una quarta part de la informació als components de cromina respecte a la luminància.

En el cas del sistema de televisió europeu, la norma 4:2:0 té les característiques de codificació següents per a cada un dels components:

	Y	Cr	Cb
Nombre d'imatges per segon	25 imatges	25	25
Nombre de línies per imatge	576 línies	288	288
Nombre de columnes per imatge	720 columnes	360	360
Nombre de bits per component	8/10/12	8/10/12	8/10/12

Taula 1. Característiques de codificació per a la norma 4:2:0 del sistema de televisió europeu

Si suposem que quantifiquem cada component amb 256 nivells (8 bits), la taxa de bits necessària per a transmetre un senyal de televisió en format 4:2:0 dependrà de l'equació següent:

$$r_b = 25 \frac{\text{imatges}}{\text{s}} \times (576 \times 720 + 2 \times 288 \times 360) \frac{\text{components}}{\text{imatge}} \times 8 \frac{\text{bits}}{\text{component}} = 124,416 \cdot 10^6 \text{ bits/s} \cdot$$

Per tant, el senyal de vídeo sense comprimir requereix un nombre de bits molt important, per la qual cosa la norma 4:2:0 (i també la 4:2:2) només s'utilitza en intercanvi de senyal entre equips de vídeo. Si continuem posant la nostra línia ADSL d'1 Mbps, necessitarem llogar un total de 125 línies per a transmetre un senyal de vídeo en temps real, d'acord amb aquesta norma. Aquest fet ens indica que abans de transmetre o emmagatzemar la informació de vídeo l'hauréu de comprimir, per a la qual cosa s'acostuma a utilitzar l'estàndard de compressió MPEG-2. La norma 4:2:0 és especialment important perquè és el format d'origen (és a dir, d'abans de comprimir) que s'utilitza en la difusió de la televisió digital, tant en la versió de cable com en les versions de satèl·lit o terrestre. També és el format que s'utilitza en el sistema DVD de vídeo abans de realitzar la compressió en MPEG-2.

4. Compressió de fonts. Necessitat i conceptes bàsics

La digitalització de senyals analògics aporta avantatges considerables entre els quals es poden esmentar la protecció davant sorolls, la facilitat per a encriptar els senyals, la possibilitat de processar les dades digitalment, etc. No obstant això, també té d'alguns inconvenients. Hem vist que, en general, la digitalització directa mitjançant PCM implica un elevat nombre de bits. En efecte, el senyal d'àudio analògic en banda base té una amplada de banda de prop de 20 kHz, però en digitalitzar-la i codificar-la en PCM es converteix en un senyal de prop de 700 kbps (hem calculat 1,4 Mbps aproximadament per a tots dos canals estereofònics). De la mateixa manera, un senyal de vídeo analògic té una amplada de banda aproximada de 5 MHz. En canvi, la digitalització de la seqüència de vídeo en un format PCM com el 4:2:0 necessita transmetre una taxa de bits de $124,4 \cdot 10^6 \text{ bit/s} = 124,4 \text{ Mbps}$.

Encara que s'usin codis i modulacions digitals de gran eficiència espectral, els recursos espectrals (amplades de banda) necessaris per a transmetre les taxes de bits associades als senyals digitalitzats seran molt superiors als que es requereixen amb els sistemes analògics.

Prenent la codificació PCM com a punt de partida, és necessari aplicar un procés de reducció de la taxa de bits que permeti la transmissió dels senyals sense una pèrdua aparent de qualitat. Tots els sistemes de compressió d'informació utilitzats actualment es poden descompondre en els tres mòduls o elements que es mostren a la figura 7, i que analitzarem amb cert detall en els apartats següents.

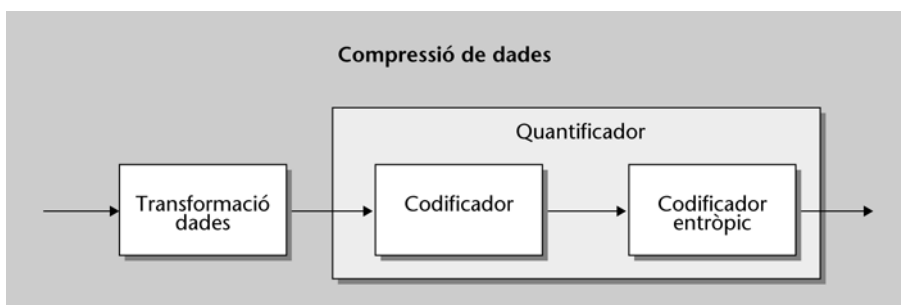


Figura 7. Elements d'un sistema de compressió de dades

4.1. La transformació de les dades

El bloc de transformació de dades és un procediment genèric que s'utilitza per a representar la informació d'una manera alternativa i en la qual la redundància en les dades originals resulta més evident. És necessari que aquesta transformació sigui invertible, és a dir, que a partir de les dades transformades

puguem recuperar exactament la informació original. Totes les transformacions utilitzades en els estàndards actuals de compressió d'àudio i vídeo són invertibles, la qual cosa significa que en aquest bloc concret no s'introdueix cap tipus de pèrdua.

Considerem com a exemple que n'apliquem una transformada de Fourier als senyals que representen la informació d'un fragment d'àudio. És possible que sigui més eficient codificar les dades resultants d'aquesta transformació que les dades originals. En efecte, en el domini transformat de Fourier és més simple identificar les regions espectrals que tenen una energia per sota del llindar d'audició, o analitzar les bandes que quedaran emmascarades per altres senyals i que, per tant, no seran audibles. Si aquesta anàlisi del senyal és factible, no tindrà sentit codificar tota la informació que el sistema auditiu humà és incapaç d'escoltar i serà més rendible dedicar els bits a les zones audibles. El resultat final és que, en aquest domini transformat, podem codificar la informació de manera més eficaç. El sistema MP3 utilitza els principis de codificació exposats per a comprimir senyals d'àudio d'alta fidelitat.

La transformació de les dades pot ser de naturalesa molt diferent (Fourier, cosinus, ondeta o *wavelet*, diferència, etc.) o, fins i tot, és possible que no sigui necessari realitzar la transformació si és que hi ha procediments eficients per a la compressió directa o els recursos computacionals de l'aplicació són escassos. En alguns casos, és possible es realitzin diverses transformacions simultàniament per a diferents components del senyal i que la codificació es realitzi prenent dades parcials de cada una. Aquest és el cas, per exemple, de l'estàndard MPEG-2 vídeo, en el qual es combinen diferents estratègies de codificació en funció del tipus d'informació i de les característiques del senyal.

4.2. Quantificador

El bloc de quantificació és el responsable principal de la introducció de les pèrdues en el sistema de compressió. Essencialment, la seva funció consisteix a aproximar els resultats obtinguts a la sortida del bloc de transformació mitjançant un conjunt finit de valors possibles (alfabet finit de resultats possibles). Aquesta aproximació realitzada sobre els resultats de la transformació impedeix que el senyal original es pugui recuperar de manera exacta. Per tant, en les aplicacions en les quals es vulgui recuperar la informació de manera exacta, el bloc de quantificació no haurà de ser present. La presència o no d'aquest bloc en l'esquema general del compressor determina una possible classificació dels mètodes de compressió.

La compressió pot ser amb pèrdues (*lossy*) o sense pèrdues (*lossless*) en funció de si la informació que es recupera coincideix exactament amb l'original o simplement és una aproximació. Els mètodes de **compressió sense pèrdues** s'utilitzen sobretot en aplicacions de codificació de dades binàries d'aplicacions

informàtiques, en les quals és absolutament necessari recuperar la informació original. Els formats més populars són el *zip*, el *rar* i l'*ar*. No obstant això, tots aquests mètodes de compressió (WinZip, WinRar) donen uns resultats extremadament ineficaços quan s'apliquen a senyals audiovisuals. És fàcil comprovar que, quan s'intenta comprimir un fitxer *bitmap* (*.bmp) o un arxiu d'àudio PCM (*.wav) mitjançant aquest tipus d'utilitats, els factors de compressió obtinguts són molt petits. Per a resoldre aquest problema, és necessari desenvolupar algoritmes sense pèrdues específics per a les aplicacions d'àudio, imatge i vídeo. Generalment, aquests sistemes específics aconseguixen obtenir factors de compressió de 2 o 3, millors que els genèrics WinZip i WinRar però insuficients per a moltes aplicacions de transmissió o arxivament de dades.

En tractament d'imatges, els mètodes de compressió sense pèrdues troben aplicació en la codificació d'imatges mèdiques o científiques en les quals la pèrdua de part de la informació pot ser crítica. El format JPEG disposa d'una versió de codificació sense pèrdues que se sol emprar en aquestes aplicacions. En àudio, també hi ha algunes aplicacions que permeten comprimir la informació i recuperar de manera exacta l'original. La seva aplicació més coneguda és la transferència d'arxius musicals, especialment de música clàssica en qualitat CD-àudio per d'Internet. El format de compressió d'àudio sense pèrdues més estès és l'APE.

La **compressió amb pèrdues** és la més habitual en la codificació de senyals de vídeo i àudio. Les pèrdues són tolerables sempre que la qualitat dels senyals descodificats sigui acceptable. El principi general sobre el qual se sosté la codificació amb pèrdues és que no resulta necessari codificar els components de la informació que no són audibles o visibles pels sistemes de percepció humana. Per tant, aquests mètodes es basen en les característiques psicofisiològiques dels sistemes auditiu i visual que són, en última instància, els que han d'avaluar la qualitat de l'algoritme de compressió. Per aquest motiu, és fonamental comprendre les limitacions i característiques d'aquests sistemes de percepció per a dissenyar codificadors en què les pèrdues d'informació siguin poc evidents o, fins i tot, inapreciables.

L'avantatge principal d'aquestes estratègies de codificació és que aconseguixen uns factors de compressió sensiblement majors que els que s'obtenen amb els mètodes sense pèrdues. A més, solen ser mètodes escalables amb l'aplicació, és a dir, el grau de pèrdua de qualitat que es tolera depèn de l'àmbit al qual es destina el codificador. Així, la qualitat que s'exigeix en una aplicació de vídeo en directe per una xarxa de telefonia mòbil és molt menor que la que s'exigeix per a la difusió de televisió. Per tant, un codificador de vídeo per a aplicacions de videoconferència en xarxes mòbils haurà de comprimir les dades molt més que un codificador per a aplicacions de televisió, ja que les taxes de transferència de bits que admeten ambdues aplicacions són clarament diferents.

Quantificació, compressió amb pèrdues, i qualitat

Considerem un esquema de compressió com el de la figura 7 en què la transformació que s'aplica a les dades és la transformada cosinus. Els resultats d'aplicar aquesta transformada depenen dels valors del senyal PCM d'entrada (nombre de bits amb què ha estat codificada) i de la precisió numèrica amb què s'ha realitzat la transformada. Un exemple típic seria el fet que els nombres obtinguts a la sortida del bloc transformador siguin enters compresos entre -8.192 i 8.191 (la transformada es realitza en coma fixa amb una precisió de 14 bits). En una primera aproximació, podem suposar que si mantenim aquesta precisió podrem recuperar de manera pràcticament exacta la informació del senyal original.

La funció del bloc quantificador és reduir el nombre total de possibles paraules que s'hauran de codificar. En el nostre exemple, cada element de la transformada pot prendre 2^{14} valors possibles. El bloc quantificador pot aproximar el resultat de la transformada, per exemple, en passos de 8, de manera que el número obtingut després de la transformació sempre s'aproxima al múltiple de 8 més proper. Així, en aquest exemple, els únics resultats possibles serien..., -24 , -16 , -8 , 0 , 8 , 16 ..., de manera que el nombre total de missatges possibles a codificar es redueix de 16.384 a $2.048 = 2^{11} = 2^{14}/2^3$.

Aquesta reducció en el nombre de possibles paraules a transmetre comportarà una disminució del nombre total de bits a enviar. Així, per exemple, ara només es necessiten 11 bits per enviar un dels resultats, mentre que originalment se'n necessitaven 14. El cost d'aquesta compressió és que els valors de la transformada han estat aproximats i la informació original només es podrà recuperar de manera aproximada.

El factor de compressió està directament relacionat amb el pas de quantificació. Estàndards com el JPEG, o les diferents versions dels MPEG àudio i vídeo, permeten que l'usuari especifiqui el factor de compressió que vol obtenir. Augmentar el factor de compressió és equivalent a augmentar el pas de quantificació. En efecte, si en el nostre exemple posem un pas de quantificació de 32, ara tots els resultats de la transformada s'hauran d'aproximar per múltiples de 32. El nombre total de resultats possibles es redueix a 512 ($= 2^9 = 2^{14}/2^5$), amb la qual cosa cada missatge es podrà codificar amb menys bits. No obstant això, hi ha una pèrdua de qualitat a conseqüència del fet que els resultats han estat aproximats amb un error major.

La relació entre el factor de compressió i la qualitat no sempre és evident. En alguns casos, es pot anar augmentant el factor de compressió sense que el resultat afecti, almenys aparentment, la qualitat. En altres casos, un petit canvi en el factor de compressió pot provocar una pèrdua molt important de qualitat. En general, el marge de treball ideal per al pas de quantificació és aquell en què apareix una relació directa entre el factor de compressió i la qualitat (*graceful degradation regions*).

4.3. Codificador entròpic

L'últim bloc d'un compressor que ens queda per discutir és el codificador entròpic. En general, el bloc de codificació consisteix a assignar una paraula de codi binari a cada un dels possibles missatges que pot generar una font. En el nostre cas, atenent la disposició dels elements a la figura 7, la font que hem de codificar és la sortida del quantificador.

En l'exemple anterior, el resultat de transformar i quantificar el senyal ens donava nombres enters, múltiples de 8 (o múltiples de 32, en el segon cas) que estaven compresos entre -8.192 i 8.191 . Per tant, teníem un total de 2^{11} (2^9 , en el segon cas) possibles missatges de manera que, en principi, semblaria natural assignar a cadascun una paraula codi d'11 bits (9 bits). Els codificadors entròpics intenten determinar les assignacions de les paraules codi d'una manera més eficient que aquesta estratègia tan directa.

La idea general d'un codificador entròpic és aprofitar l'estadística amb què es produeixen els missatges. Els codificadors entròpics més utilitzats són els

codis de longitud variable en què el nombre de bits d'una paraula codi depèn de les probabilitats amb què es produeixen els missatges. La idea bàsica és assignar paraules de **poca longitud** (pocs bits) als missatges que tenen una **major probabilitat**, i paraules de més longitud als missatges que es produeixen amb menys freqüència. Aquest argument sembla molt intuïtiu, ja que assignem pocs bits als missatges que ocorren molt sovint i que, per tant, s'espera que aportin poca informació. En canvi, els missatges que es produeixen amb molt poca freqüència tindran assignades paraules codi amb més bits, la qual cosa correspon a més informació.

En general, podem afirmar que l'estadística dels missatges d'una font es modifica quan s'apliquen transformacions als senyals, la qual cosa ens suggereix que concentrem els esforços en la cerca de les transformacions que permetin obtenir una estadística de missatges que es pugui codificar amb un nombre reduït de bits.

En els apartats següents, estudiarem amb detall la relació entre la informació aportada per un missatge i la seva probabilitat. També establirem el límit teòric de bits amb què es pot codificar una font digital i veurem algunes estratègies per a la construcció de codis de longitud variable.

Les tarifes telefòniques i els codis de longitud variable

Considerem un exemple senzill sobre tarifes telefòniques que ens servirà per a il·lustrar de manera intuïtiva les idees bàsiques dels codis de longitud variable.

La companyia telefònica X ofereix una tarifa plana per poder trucar a 256 números amb cost reduït de 0,20 €/min. La companyia telefònica Y contraataca amb una altra oferta per a trucar a 256 números a un cost de 0,35 €/min, però permet seleccionar un total de quatre números amb un cost de 0,10 €/min.

Per a saber quina companyia ens convé més, necessitem realitzar un estudi de les nostres "estadístiques" de trucada. Podrem comprovar que si tenim 256 amics que truquem amb la mateixa freqüència resulta més rendible la tarifa de la companyia X. En canvi, si les nostres trucades se centren, en un 90%, en la parella, l'amic i l'amiga íntima, i en la nostra mare, serà més rendible utilitzar la tarifa de la companyia Y.

Problema 1

Determineu el cost de realitzar cent trucades d'un minut amb les dues tarifes anteriors sabent que la nostra "estadística" de trucades se centra, en un 90%, en tres números, i que la resta de números és equiprobable.

Solució

Si utilitzem la companyia X el cost és el següent:

$$100 \text{ min} \times 0,2 \text{ €/min} = 20 \text{ €}$$

En canvi, amb la companyia Y, el nostre cost serà el següent:

$$90 \text{ min} \times 0,1 \text{ €/min} + 10 \text{ min} \times 0,35 \text{ €/min} = 12,5 \text{ €/min}$$

Problema 2

Determineu quin és el percentatge mínim de trucades a números seleccionats a partir del qual resulta rendible la tarifa de la companyia Y.

En els codis de longitud variable...

... (VLC, *Variable Length Codes*) els missatges amb molta probabilitat tenen assignades paraules codi de poca longitud, mentre que els missatges amb una probabilitat baixa tenen paraules codi de molta longitud.

Solució

Per a determinar el percentatge crític, hem de determinar el valor en el qual totes dues tarifes tindran el mateix cost:

$$A \text{ min} \times 0,2\text{€}/\text{min} = A \text{ min} \times P \times 0,1 \text{ €}/\text{min} + A \text{ min} \times (1 - P) \times 0,35\text{€}/\text{min}$$

Aclarint el valor de P , obtenim

$$P = 0,6 \text{ (60 \%)}.$$

5. Mesura de la informació

En aquest apartat ens centrarem en els conceptes essencials de la teoria de la informació: la mesura de la informació associada a un missatge, i l'entropia associada a la font. Els conceptes esmentats van ser desenvolupats per Claude E. Shannon.

Claude E. Shannon (1916-2001) va ser un gran matemàtic i enginyer que s'ha de considerar com el pare de la teoria de la informació. La passió pels enginyers mecànics i la ciència la hi va inculcar el seu avi, un granger que en el seu temps lliure construïa màquines per a l'agricultura i que va ser l'inventor de la rentadora. Des de molt jove, Shannon va mostrar una aptitud excel·lent per a les matemàtiques i per a l'enginyeria, i es va graduar en totes dues disciplines en la Universitat de Michigan i, posteriorment, va ampliar els seus estudis en el MIT (Massachusetts Institute of Technology), on va col·laborar amb Norbert Wiener i Vannevar Bush. Aquest últim era l'inventor de l'analitzador diferencial, un aparell calculador analògic que permetia resoldre equacions complexes mitjançant l'acció d'un elevat nombre d'elements mecànics que incloïen molles, engranatges, rodes dentades, etc. La solució d'una equació es trobava mesurant els canvis de posició que havien experimentat les diferents parts de la màquina. El procés de "programació" de la màquina per a afrontar un nou problema exigia desmuntar i tornar a muntar diverses peces segons una configuració mecànica adequada, tasca que podia durar diversos dies. Shannon va finançar els seus estudis com a "programador" de l'analitzador diferencial, i la seva funció principal era preparar-lo perquè altres científics poguessin resoldre les seves equacions.

Paral·lelament, va realitzar els seus estudis centrant-se en l'anàlisi dels relés de l'analitzador diferencial. Aquest estudi va ser l'eix principal de la seva tesi de màster en què aplica, per primera vegada, l'àlgebra de Boole a les màquines computadores, identificant la màquina de computació amb la lògica de Boole. El treball conclou amb l'article "A Symbolic Analysis of Relay and Switching Circuits" (1938), que hauria de tenir una gran influència en el desenvolupament posterior de sistemes i màquines de processament d'informació. Va realitzar la tesi doctoral aplicant conceptes d'àlgebra en genètica teòrica.

Va treballar durant molts anys en els Laboratoris Bell, on va col·laborar amb altres matemàtics i científics en aspectes clau de les comunicacions, millorant l'eficiència i la fiabilitat de les transmissions telefòniques i telegràfiques a llarga distància. Els seus treballs sobre criptologia van constituir la base del que es considera com el seu article més important: "A Mathematical Theory of Communications" (1945) en què s'introdueixen els conceptes d'entropia, mesura de la informació, capacitat d'un canal, distorsió i codis redundants, per a la protecció davant errors.

El 1950 publica un altre dels seus articles clàssics: "Programming a Computer for Playing Chess". Aquest va ser, durant molts anys, una referència en la programació d'autòmats i ordinadors per al joc dels escacs.

La passió per l'enginyeria mecànica i la computació el van mantenir ocupat durant gran part de la seva vida en el disseny de màquines que ell mateix qualificava d'"inútils". Per posar només alguns exemples, va construir una màquina calculadora en xifres romanes, unes tortugues autòmats que caminaven soles per casa seva, una màquina que realitzava malabarismes amb tres pilotes, una màquina que permetia resoldre el cub de Rubik, un ratolí automàtic que "aprenia" tècniques per a sortir de laberints, etc. La seva gran passió va ser el monocicle, en el qual passava moltes hores practicant malabarismes.



Figura 8. Claude E. Shannon.

5.1. Informació associada a un missatge

Una font de missatges es caracteritza pel nombre de missatges que pot produir i per la probabilitat amb què es produeix cada missatge. A la figura 9, es repre-

senta la idea bàsica de la caracterització d'una font de missatges en la qual es defineixen els diferents missatges m_k i les seves probabilitats associades p_k . En la pràctica, és difícil conèixer *a priori* la probabilitat amb què es produeix cada un dels missatges. No obstant això, sempre es pot realitzar un estudi en el qual es determinin les freqüències d'aparició de cada missatge i estimar la seva probabilitat a partir d'aquests resultats.

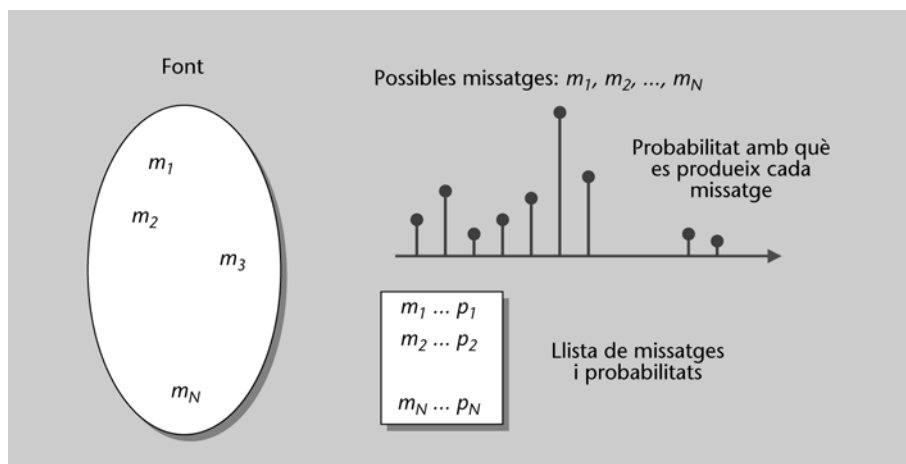


Figura 9. Caracterització d'una font de missatges

Un exemple de font de missatges...

... podria ser una imatge en blanc i negre. Els possibles missatges que genera aquesta font són el nivell de gris dels píxels. Per tant, en el cas més habitual, tindrem un total de 256 missatges possibles. Podem estimar la probabilitat de cada missatge determinant el percentatge de píxels que tenen un determinat nivell respecte al total de píxels de la imatge.

Una vegada especificada o estimada l'estadística de la font, ens haurem de plantejar establir una mesura de la quantitat d'informació associada a cada missatge. La mesura ha de verificar certs requisits que semblen naturals i que, generalment, estan associats al concepte qualitatiu d'informació. Aquests postulats són els següents:

1) **La informació associada a un missatge és positiva.** És lògic suposar que sempre que es produeix un dels possibles missatges, aquest aporta alguna mena d'informació. No té sentit que es produeixin pèrdues d'informació si rebem noves dades.

2) **Els missatges amb poca probabilitat aporten més informació que els missatges amb molta probabilitat.** Els missatges que es produeixen freqüentment ens aporten molt poca informació perquè són esperats per endavant. Així, quan se'ns informa que hi ha circulació lenta en els accessos d'entrada a Barcelona entre les 8 i les 9 del matí, som davant d'un missatge que ens aporta molt poca informació. En canvi, la notícia d'una gran nevada a Barcelona apareixerà sempre amb grans titulars als diaris ja que, per la seva baixa probabilitat, ens aportarà molta informació.

3) **La informació que aporten dos missatges simultanis estadísticament independents és la mateixa que la que aporten quan es produeixen separatament.** Això significa que la informació que ens aporten dos esdeveniments independents que es produeixen simultàniament coincideix amb

la suma de les informacions que ens aportarien si es produïssin separatament.

Podem formular aquests postulats mitjançant les equacions següents:

$$I(m_k) \geq 0 \quad (1)$$

$$I(m_i) > I(m_j) \quad \text{si} \quad p_i < p_j \quad (2)$$

$$I(m_i, m_j) = I(m_i) + I(m_j) \quad \text{si} \quad P(m_i, m_j) = P(m_i) \cdot P(m_j) \quad (3)$$

Segons teorema fonamental del càlcul, únicament existeix una funció que verifiqui les tres propietats simultàniament. D'aquesta manera, la mesura quantitativa de la informació és totalment axiomàtica, ja que es troba directament a partir de les propietats que volem que verifiqui.

Així, doncs, la informació associada al missatge m_k estarà determinada per l'única funció que verifica les tres propietats anteriors:

$$I(m_k) = -\log_b p_k \quad (4)$$

És fàcil comprovar que la definició anterior verifica les propietats per a qualsevol base b del logaritme. Quan la base del logaritme és 2, la unitat d'informació s'anomena **bit**. Si la base del logaritme és el nombre d'Euler e la unitat d'informació es mesura en **nats**.

Considerem com a exemple la informació que proporciona una moneda en caure. Hi ha dos possibles missatges (cara i creu), tots dos amb la mateixa probabilitat (1/2). La informació associada al missatge cara serà

$$I(\text{cara}) = -\log_2(1/2) = 1 \text{ bit}$$

Anàlogament, si considerem una font que pot transmetre vuit missatges equiprobables, la informació associada a cada un d'aquests missatges depèn de

$$I(m_i) = -\log_2(1/8) = \log_2 2^3 = 3 \text{ bit}$$

D'acord amb aquests resultats, per a estimar la quantitat d'informació lligada a un dels missatges de la font és necessari establir, en primer lloc, les característiques estadístiques de la font de missatges. Aquestes característiques es poden estimar de manera experimental deixant que la font generi missatges i aproximant la seva probabilitat per la freqüència relativa d'aquests. En general, si l'estimació es realitza sobre una mostra prou alta, aquest procediment proporciona resultats satisfactoris.

5.2. Entropia d'una font de missatges

Coneixent la informació associada a cada missatge, podem determinar la informació mitjana que proporciona la font. Per a això, hem de ponderar la informació de cada missatge per la probabilitat que es produeixi:

$$H = \sum_{k=1}^N p_k \cdot I(m_k) = -\sum_{k=1}^N p_k \cdot \log_2 p_k \quad (5)$$

Aquesta funció es coneix amb el nom d'**entropia** i estableix un límit al nombre de bits mitjà amb què es podran codificar els missatges d'una font. Cap codificador no podrà obtenir codis amb un nombre de bits mitjà inferior a l'entropia de la font. El nom d'*entropia* de la funció anterior es deu al fet que la seva fórmula s'assembla molt a l'entropia termodinàmica d'L. Boltzman. Recentment, algunes teories físiques han relacionat totes dues entropies (J. D. Bekenstein).

Missatges equiprobables i missatges amb probabilitats dispars

Comprovarem amb un exemple que l'entropia d'una font disminueix quan un o uns quants missatges tenen una probabilitat d'ocurrència molt més alta que els altres. Aquest resultat es pot generalitzar (vegeu el problema 3) per a demostrar que l'entropia d'una font és màxima quan tots els missatges que produeix són equiprobables.

Imaginem dues fonts F_1 i F_2 que poden produir un total de 256 missatges cada una. En la font F_1 , tots els missatges són equiprobables ($p_k = 1/256$), de manera que la informació associada a cada missatge és de 8 bits i l'entropia de la font també serà de 8 bits.

$$H(F_1) = \sum_{n=1}^{256} \frac{1}{256} \cdot \log_2 \left(\frac{1}{256} \right) = 8 \text{ bits}$$

La font F_2 té un missatge amb una probabilitat molt alta, i la resta de missatges tenen la mateixa probabilitat:

$$\begin{aligned} m_1 &\rightarrow p(m_1) = p_1 \\ m_k &\rightarrow p(m_k) = \frac{(1-p_1)}{255} \quad k = 2, \dots, 256 \end{aligned}$$

En aquest cas, l'entropia depèn de

$$H(F_2) = -p_1 \cdot \log_2 p_1 - (1-p_1) \cdot \log_2 \left(\frac{1-p_1}{255} \right). \quad (6)$$

En el gràfic de la figura 10, es mostra el valor de l'entropia d' F_2 en funció del valor de la probabilitat p_1 . Podem veure que, a mesura que la probabilitat p_1 s'apropa a la unitat, l'entropia de la font disminueix. Això coincideix amb la idea intuïtiva que, a mesura que p_1 augmenta, el resultat que proporcionarà la font és més previsible (el missatge m_1 té una probabilitat d'ocurrència més alta que la resta), de manera que la informació mitjana que proporcionarà serà menor. En el cas límit en què la probabilitat p_1 sigui la unitat, la font no proporcionarà cap informació. Un altre aspecte interessant d'observar és que el valor màxim de l'entropia es produeix quan $p_1 = 1/256$. En efecte, encara que a la figura 10 pugui donar la sensació que el màxim de l'entropia és en l'origen, substituint en l'equació (6) $p_1 = 0$ i $p_1 = 1/256$ obtenim, respectivament 7,9943 bits i 8 bits. Aquest resultat també sembla intuïtiu, ja que quan tots els símbols són equiprobables, la nostra incertesa sobre quin serà el missatge que produirà una font és màxima, d'aquesta manera és natural que la informació mitjana proporcionada per la font també prengui el seu valor màxim. Demostrarem aquest resultat matemàticament en el problema 3.

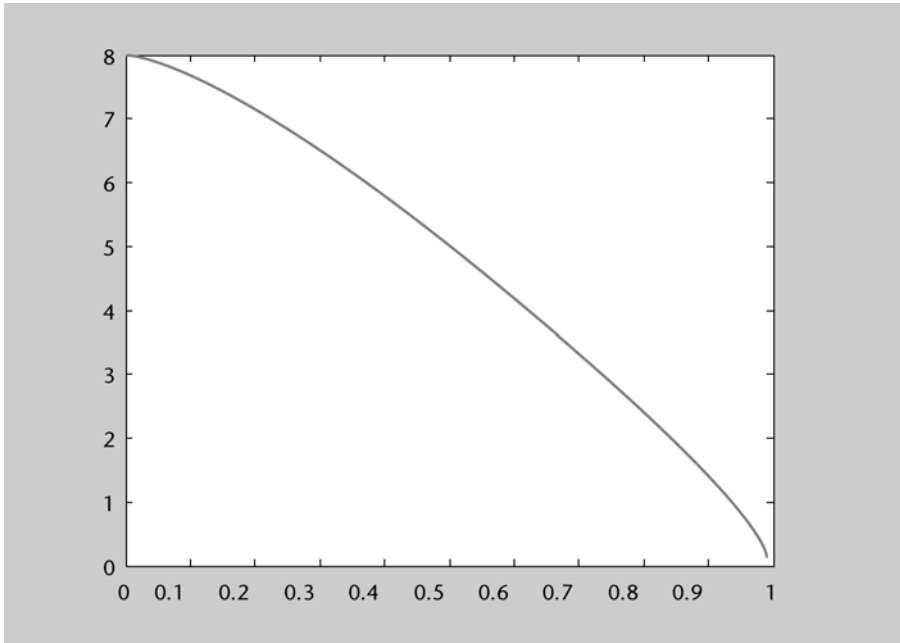


Figura 10. Evolució de l'entropia amb la probabilitat d'un dels missatges (resta de missatges equiprobables)

Problema 3

Demostreu que, per a una font de M missatges, l'entropia és màxima quan tots els missatges són equiprobables.

Solució

Quan els missatges són equiprobables l'entropia depèn de

$$H = \sum_{i=1}^M \frac{1}{M} \cdot \log_2(M) = \log_2(M) \cdot$$

Per tant, hem de demostrar que, en un cas general,

$$H(F_M) \leq \log_2 M,$$

en què F_M indica una font genèrica d' M missatges.

Considerem la diferència següent:

$$H(F_M) - \log_2(M) = \sum_{i=1}^M p_i \cdot \log_2\left(\frac{1}{p_i}\right) - \sum_{i=1}^M p_i \cdot \log_2(M) = \sum_{i=1}^M p_i \cdot \log_2\left(\frac{1}{Mp_i}\right)$$

I utilitzem la desigualtat següent:

$$\ln y \leq y - 1 \Rightarrow \log_2 y \leq (y - 1) \cdot \log_2 e$$

Finalment obtenim

$$H(F_M) - \log_2(M) \leq \log_2 e \cdot \left(\sum_{i=1}^M p_i \left(\frac{1}{Mp_i} - 1 \right) \right) = \log_2 e \cdot \left[\sum_{i=1}^M \frac{1}{M} - \sum_{i=1}^M p_i \right] = 0.$$

La qual cosa demostra la proposició.

6. Codis de longitud variable. Conceptes bàsics

Tal com hem explicat més amunt, la idea bàsica dels codis de longitud variable és assignar paraules codi de longituds diferents en funció de la probabilitat dels missatges. Els missatges més probables es codificaran amb paraules amb un nombre més petit de bits que els missatges menys probables.

Considerem com a exemple els missatges de la font que es resumeixen en la taula 2, en la qual els codis assignats a cada missatge s'han triat perquè compleixin una sèrie de restriccions que detallarem més endavant.

missatges	probabilitat	codi
m_1	0,4	0
m_2	0,3	10
m_3	0,1	1100
m_4	0,1	1101
m_5	0,1	1110

Taula 2. Exemple de codis de longitud variable

Podem calcular el nombre mitjà de bits que utilitza el codi proposat ponderant el nombre de bits que s'utilitza per a cada missatge amb la probabilitat que es produeixi el missatge.

$$N_{\text{mitjà}} = \sum_{k=1}^N p_k \cdot N_k = 1 \cdot 0,4 + 2 \cdot 0,3 + 4 \cdot 0,3 = 2,2 \text{ bits}$$

El nombre de bits mitjà ha de ser superior a l'entropia. En efecte, l'entropia representa la informació mitjana de la font, per la qual cosa qualsevol codi pràctic que utilitzem tindrà un nombre de bits mitjà major o igual que l'entropia. En el nostre cas,

$$H = -\sum_{k=1}^N p_k \cdot \log_2 p_k = 2,04 \text{ bits}$$

Problema 4

Determineu, utilitzant la taula 2, la seqüència de bits amb què quedaria codificada la seqüència de missatges següent: $m_3 m_5 m_1 m_1 m_3 m_2 m_1 m_1$.

Solució

Per a obtenir la seqüència, de bits simplement hem de substituir cada missatge pel codi corresponent. En el nostre exemple obtenim:

1100 1110 0 0 1100 10 0 0

Problema 5

Imaginem una font amb cinc missatges que ha estat codificada amb la taula 2. Determineu la seqüència de missatges a què correspon la seqüència de bits següent:

010011010110010

Solució

La manera més usual de determinar la seqüència de missatges consisteix a anar descodificant paquets de bits una vegada es troben en la taula. Si un paquet de bits no es troba en la taula, s'augmenta el paquet prenent un bit addicional. Així:

Paquet 1:	0	Correspon al missatge m_1 .
Paquet 2:	1	No correspon a cap missatge, augmentem en paquet amb el bit següent.
Paquet 2:	10	Correspon al missatge m_2 .
Paquet 3:	0	Missatge m_1 .
Paquet 4:	1	No correspon a cap missatge.
Paquet 4:	11	No correspon a cap missatge.
...		
Paquet 4:	1101	Correspon al missatge m_4 .

Utilitzant aquest procediment, la seqüència de missatges final és

$m_1 m_2 m_1 m_4 m_1 m_3 m_2$.

El codi que hem utilitzat en aquests exemples ha estat proposat per endavant. Hem vist que el seu nombre de bits mitjà està molt pròxim a l'entropia, i que és senzill obtenir seqüències de bits a partir d'una seqüència de missatges i viceversa. Tanmateix, hauríem de ser conscients que no tots els codis tindran aquestes dues propietats i que hauríem de trobar un procediment sistemàtic que ens permetés generar aquest tipus de codis. En l'apartat següent, examinarem amb detall la construcció i algunes propietats dels codis de Huffman. Aquests codis i les seves variants s'utilitzen en la majoria de compressors de senyals d'àudio i vídeo (MPEG-1, MPEG-2, MPEG-4, MP3, etc.).

7. Codis de Huffman

En aquest apartat estudiarem els codis de Huffman, els quals s'han convertit en una de les tècniques més utilitzades en les tecnologies de la comunicació digital. Aquests codis deuen el seu nom a David A. Huffman.

David A. Huffman (1925-1999) ha estat un dels científics que han tingut més influència en les tecnologies de la informació, aportant contribucions significatives en diferents camps com la teoria de màquines d'estat finit, les metodologies de disseny de circuits lògics, el disseny de senyals per a radar i, sobretot, per la seva aportació a la compressió de senyals mitjançant els codis de longitud variable que porten el seu nom. Actualment, els codis de Huffman o algunes de les seves múltiples variants s'utilitzen en pràcticament totes les aplicacions que requereixen la compressió i transmissió de dades digitals: mòdems, faxes, xarxes d'ordinadors, arxius d'àudio MP3, arxius de vídeo (MPEG-2, DIVX), etc.

El 1951 David A. Huffman era un estudiant de l'assignatura de Teoria de la informació, impartida pel Professor Robert M. Fano en el MIT (Massachusetts Institute of Technology). Durant aquest curs, el professor Fano va oferir als seus estudiants la possibilitat de realitzar un treball amb el qual podien quedar exempts de presentar-se a l'examen de final de curs. El treball consistia a determinar els codis més eficients per a representar nombres, lletres o altres símbols (missatges) utilitzant codis binaris. El que el professor Fano no va esmentar era que tant Claude E. Shannon com ell mateix feia diversos anys que col·laboraven i intentaven obtenir un mètode òptim per a l'obtenció d'aquests codis binaris sense que, de moment, haguessin tingut cap èxit. Huffman va treballar en el problema durant uns quants mesos, creant diferents alternatives per a l'obtenció de codis sistemàtics sense poder arribar a demostrar que alguna de les solucions era òptima. Finalment, desanimat, i quan ja començava a estudiar per presentar-se a l'examen final, va trobar la demostració definitiva.

Actualment, els codis de Huffman constitueixen una d'aquestes poques idees fonamentals que s'utilitzen en pràcticament totes les tecnologies de comunicació de dades digitals. Una altra de les contribucions especialment importants de David A. Huffman, i de la que ell se sentia particularment orgullós, va ser la primera tècnica de metodologia formal per al disseny de circuits lògics seqüencials que va desenvolupar en la seva tesi doctoral.

Encara que els codis de Huffman han ajudat molts inversors i tecnòlegs a fer-se milionaris, l'únic profit que Huffman va obtenir pels codis va ser la dispensa en l'examen final de Teoria de la informació. Huffman mai no va tractar d'obtenir una patent per alguna de les seves contribucions, encara que és evident el substanciós que haurien estat per a ell aquestes patents, especialment la dels codis. Per acabar, direm que Huffman va ser deixeble directe de Shannon, ja que aquest va passar diverses hores ensenyant-lo a muntar en monocicle, fins que finalment també es va convertir en un expert.

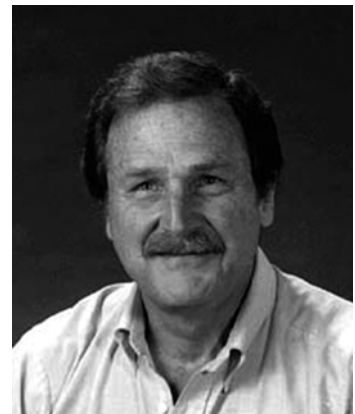


Figura 11. David A. Huffman.

7.1. Obtenció sistemàtica dels codis de Huffman

Avui dia, els codis de Huffman (1951) representen la tècnica més utilitzada per a l'eliminació de la redundància estadística en la codificació. El procediment per a obtenir aquests codis és conceptualment simple, encara que, com veurem, des del punt de vista computacional resulta complex. És possible demostrar que els codis de Huffman són òptims en el sentit que aconseguen el nombre més petit de símbols de codi per símbol de font amb la restricció que els missatges s'han de codificar un a un (cada missatge té associada una única paraula codi).

També es pot demostrar que el nombre de bits mitjà que s'aconsegueix obtenir amb un codi de Huffman sempre està comprès entre l'entropia de la font i l'entropia més un:

$$H < \bar{N}_{\text{Huffman}} < H + 1 \quad (7)$$

El procediment per a obtenir un codi de Huffman es pot resumir en les etapes següents:

- 1) Ordenar els missatges per probabilitats decreixents.
- 2) Crear un arbre combinant sempre els dos nodes amb menor probabilitat. El resultat de la combinació dels dos nodes és un nou node la probabilitat del qual és la suma de les probabilitats dels nodes que l'han construït. Continuar amb aquest procediment fins a obtenir un únic node.
- 3) Per a assignar els codis als missatges, n'hi ha prou de recórrer l'arbre des de l'arrel fins a cada un dels missatges, assignant un 0 o un 1 en funció de si prenem una branca cap a dalt o cap a baix (l'assignació de zeros i uns a les branques pot ser arbitrària).

El procediment anterior és bastant més clar si s'il·lustra amb un exemple senzill com el que es mostra a la figura 12. Veiem com els nodes es van combinant de dos en dos, triant sempre els de probabilitat més petita. Si hi ha diversos nodes amb la mateixa probabilitat, se'n poden combinar dos qualssevol. La probabilitat assignada al node arrel sempre haurà de ser la unitat.

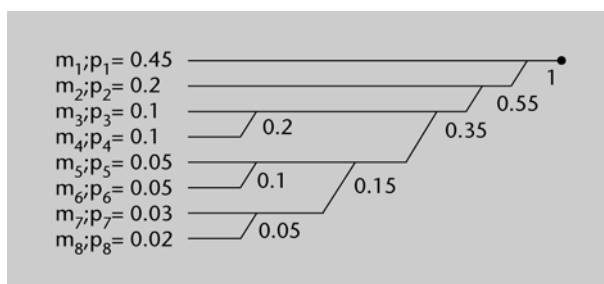


Figura 12. Procés de creació del codi de Huffman

La figura 13 mostra el mateix arbre, però ara s'utilitza per a l'assignació de codis als missatges. Així, per a determinar el codi que assignem al missatge m_4 haurem de partir del node arrel, fer dos passos cap a baix i després dos passos més cap a les branques superiors. El codi resultant per a aquest missatge és 1100. Fixeu-vos que, si seguim aquesta construcció de l'arbre de Huffman, el missatge amb menys probabilitat sempre queda codificat amb tots els símbols igual a la unitat.

Una vegada determinat el codi de Huffman per a aquesta font, podem calcular el nombre mitjà de bits que s'utilitzen i obtenir un total de 2,4 bits. Si calculem l'entropia de la font original obtindrem 2,344 bits, la qual cosa dóna una idea de la qualitat del codi.

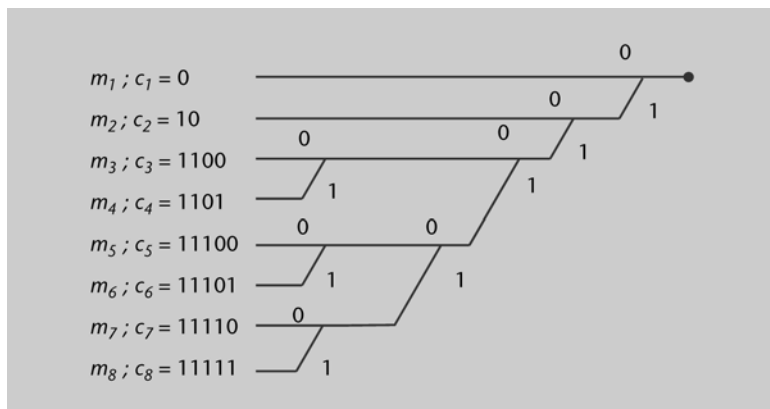


Figura 13. Procés de generació del codi a partir de l'arbre de Huffman

Una de les característiques dels codis de Huffman és que permeten descodificar la seqüència de dades a partir de la concatenació directa dels codis i sense necessitat de “senyalitzar” el principi i el final d'un missatge. Pot ser que això no sembli trivial, ja que els codis són de longitud variable, per la qual cosa és possible que no sigui evident determinar quan comença i quan acaba cada codi. No obstant això, la descodificació és possible i a més és senzilla perquè, tal com està construït l'arbre de Huffman, cap dels codis llargs no té com a subcadena una paraula codi més curta. Tots els codis que compleixen que una paraula codi mai no és el prefix de cap altra paraula codi es poden descodificar sense ambigüitat. Aquests codis s'anomenen *de prefix*, i es pot demostrar que tots compleixen l'equació (7). Els codis de Huffman són els codis de prefix que tenen un nombre de bits mitjà més reduït.

Per il·lustrar el procediment de descodificació, considerem la cadena de bits 11101110100101111..., el procés de la qual es representa esquemàticament a la figura 14. La descodificació començaria comprovant si el primer bit (1) és una paraula codi, buscant-lo en la taula de paraules codi amb un únic bit. Com que no és així, comprovaríem si els dos primers bits (11) poden ser una paraula codi. Continuaríem amb aquest procediment fins que trobéssim la primera paraula codi 11101, que sí que es troba en la taula, i que correspon al missatge m_6 . Procedint d'aquesta manera, arribaríem a descodificar la seqüència de missatges completa que es correspon amb els codis c_6, c_4, c_1, c_1, c_2 i c_8 .

Un dels aspectes més crítics en els codis de Huffman i, en general, en els codis de longitud variable, és l'aparició d'errors en el canal. En efecte, quan sorgeix algun error de bit en el canal, es pot produir un canvi de paraula codi que comporta la pèrdua del sincronisme i la possibilitat que apareguin un gran nombre d'errors consecutius. Algunes vegades, el receptor pot detectar automàticament l'aparició dels errors perquè no troba les paraules codi en les taules (o l'estadística

d'aparició no es correspon amb l'esperada). Encara que hi ha algoritmes perquè el descodificador es pugui resincronitzar amb la cadena de dades després d'un error, aquests algoritmes són complexos, i el seu cost computacional, molt elevat, per la qual cosa s'usen molt poc en aplicacions pràctiques.

Taula	Seqüència de bits
$c_1 = 0$	111011101001011111...
$c_2 = 10$	1 1 1 0 1 → $c_6 = 11101$
$c_3 = 1100$	111011101001011111...
$c_4 = 1101$	1 1 0 1 → $c_4 = 1101$
$c_5 = 11100$	111011101001011111...
$c_6 = 11101$	0 → $c_7 = 0$
$c_7 = 11110$	
$c_8 = 11111$	

Figura 14. Exemple de descodificació d'una cadena de missatges

La solució més habitual per protegir-se dels possibles errors de bits en el canal consisteix a introduir **paraules de resincronització** amb una freqüència determinada (paraula única, amb el valor i longitud coneguts pel descodificador i que serveix per a buscar un punt de resincronització quan s'han produït errors). Les paraules de resincronització permeten utilitzar algoritmes molt simples de descodificació: detectar pèrdua de sincronisme (errors) i buscar la paraula única que ens permetrà tornar-nos a sincronitzar. La freqüència amb què s'introdueix la paraula de resincronització depèn de la probabilitat amb què es produeixin errors. Si els errors són molt probables, convé introduir-la amb més freqüència. No obstant això, és evident que introduir la paraula de resincronització amb molta freqüència redueix l'eficiència del compressor. Els estàndards de compressió MPEG per a senyals d'àudio i vídeo utilitzen aquesta tècnica per a recuperar el sincronisme del descodificador si s'han produït errors. La freqüència amb què s'insereix la paraula de resincronització en MPEG depèn de l'aplicació. Així, en difusió via satèl·lit, la paraula s'insereix moltes més vegades que en difusió per cable, ja que en aquest últim cas els senyals estan més protegits contra errors i interferències.

L'obtenció del codi de Huffman una vegada es disposa de l'estadística del senyal és conceptualment molt senzilla. No obstant això, els algoritmes associats no són trivials i requereixen realitzar força simplificacions i comparacions en la taula de missatges. Un altre dels problemes importants és l'estimació de les estadístiques amb què es produeixen els missatges. Les estimacions dels missatges menys probables són poc fiables i s'ha treballat amb mostres molt grans per a obtenir-los. Sovint, s'obté que la probabilitat estimada d'alguns missatges és nul·la, ja que no s'han produït cap vegada en la seqüència que s'ha utilitzat per a determinar les estadístiques. Encara que el procediment de Huffman es pot aplicar de la mateixa manera en aquests casos, és possible que

els valors de la taxa de bits mitjana amb què es codifica la font no siguin tan òptims com es podria esperar. Hi ha diverses alternatives per a simplificar el disseny de la taula de codificació de Huffman. Totes són solucions subòptimes que simplifiquen l'anàlisi de la font i la construcció de la taula de codificació. Un exemple d'aquestes solucions subòptimes són els codis de Huffman truncats que exposarem més endavant.

L'estudiant pot trobar la descripció dels codis de Huffman truncats en el subapartat 7.2 d'aquest mateix mòdul.

Uns codis de longitud variable especialment interessants són els **codis reversibles** (*reversible variable length codes*), que s'utilitzen en codificació de vídeo en l'estàndard MPEG-4. Aquests codis tenen la propietat que es poden descodificar quan es llegeixen en el sentit invers a l'habitual (*backward*). D'aquesta manera, quan es produeix un error en el canal, hem d'esperar fins a la pròxima paraula de resincronització i llegir en sentit invers, amb la qual cosa serem capaços de descodificar part del missatge. Els codis reversibles són menys eficients que els codis de Huffman convencionals quant al nombre mitjà de bits amb què codifiquen la font. La figura 15 mostra gràficament la idea dels codis reversibles.

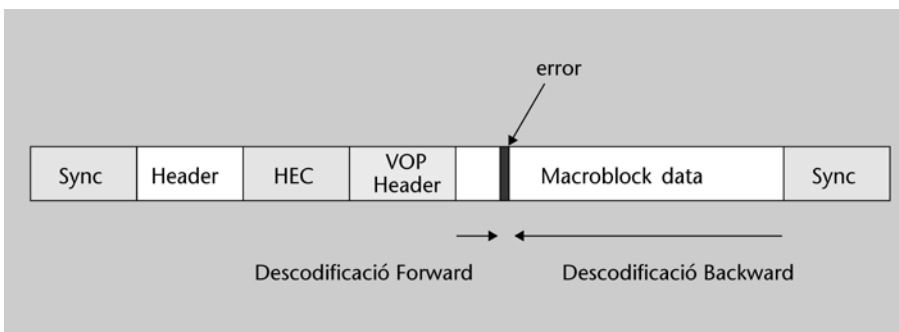


Figura 15. Codis de longitud variable reversibles

En aplicacions d'arxivament de dades en fitxers, és recomanable que la codificació i la descodificació es realitzin emprant memòries intermèdies (*buffers*) temporals amb una longitud que sigui un múltiple de 8 bits. Així, quan la memòria intermèdia és plena, es pot emmagatzemar la seqüència de bits, aprofitant una paraula d'un byte o múltiple del byte.

7.2. Codis de Huffman truncats

Una de les variants més comunes del mètode de Huffman són els codis truncats, que es basen a descartar els missatges menys probables del procés d'anàlisi del codi de Huffman. Aquests missatges es codifiquen posteriorment utilitzant paraules de longitud fixa precedides d'un prefix proporcionat per la mateixa anàlisi de la font simplificada. Les etapes de l'anàlisi del codi es resumeixen en els punts següents:

1) S'ordenen els N missatges de la font per ordre de probabilitat decreixent.

2) Es trien els M missatges menys probables i es genera un nou missatge L que representa els $(N - M)$ missatges descartats. La probabilitat d' L és la suma de les probabilitats dels missatges amb baixa probabilitat que representa.

3) Es construeix el codi de Huffman associat a la nova font de $M + 1$ missatges.

4) Es construeix el codi dels $N - M$ missatges de menor probabilitat, emprant com a prefix el codi de Huffman assignat al nou missatge L , seguit d'un codi binari convencional de longitud constant.

El procediment es resumeix esquemàticament a la figura 16, on hem suposat que es descarten els cinc missatges amb menys probabilitat. La suma de les probabilitats d'aquests missatges s'insereix en la taula de missatges ordenats prèviament (en el nostre exemple, és el cinquè missatge més probable) i es construeix el codi de Huffman amb aquesta nova estadística. El codi resultant s'utilitza com a prefix per a assignar les paraules que representen els missatges originals. En aquest exemple, hem d'afegir tres bits addicionals per a poder representar els cinc missatges.

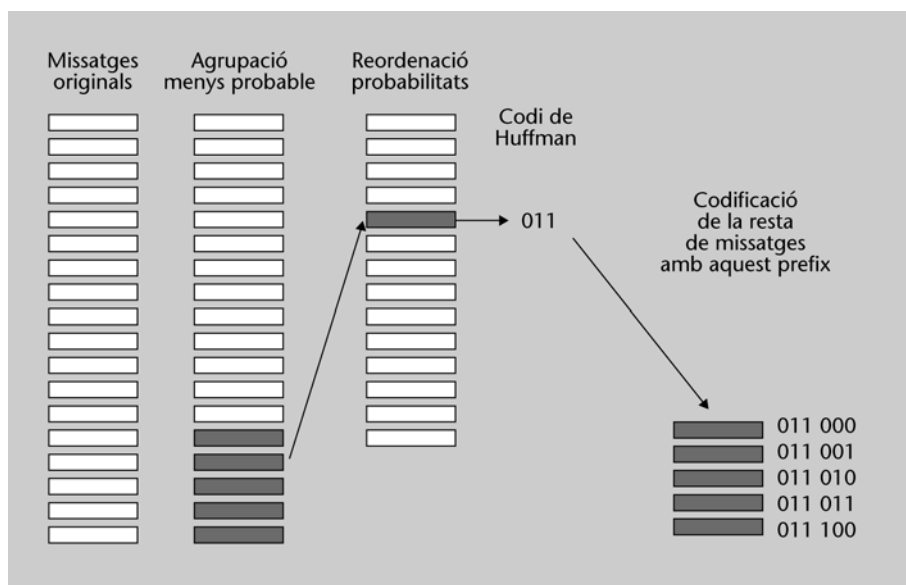


Figura 16. Construcció dels codis de Huffman truncats

8. Transformacions de dades

Hem vist que els codis de longitud variable obtenen una reducció eficient del nombre de bits mitjà d'una font sempre que hi hagi diferències importants entre les probabilitats amb què es produeixen els missatges. És per això que és evident que un dels aspectes més importants en el disseny de qualsevol compressor consisteix a utilitzar transformacions de la font de missatges perquè produeixin uns missatges amb molta més probabilitat que d'altres. Generalment, en la codificació de senyals d'àudio i vídeo, s'utilitzen diversos tipus de transformacions –diferencial, de predicció, transformada cosinus, etc.– depenent de les característiques dels components que es codifiquen. Les transformacions concretes depenen de l'aplicació, fins i tot en alguns casos es poden combinar diversos mètodes. Així, per exemple, en els casos en què els terminals tinguin restriccions de cost computacional, les transformacions seran més simples i segurament menys eficients des del punt de vista de compressió. Per il·lustrar la idea general, considerarem amb cert detall la transformació diferencial, descriurem qualitativament els mètodes predictius, i conclourem amb alguns exemples molt simplificats de la codificació de veu en telefonia mòbil GSM (*global system to mobile communication*).

L'estudi detallat de les transformacions concretes utilitzades en cada un dels sistemes audiovisuals escapa als propòsits d'aquest mòdul.

8.1. Codificació diferencial. Aplicació en la codificació d'imatge

L'esquema bàsic d'un codificador diferencial es representa a la figura 17. Com a exemple, el senyal d'entrada podria ser la luminància d'una línia de la imatge. La transformació utilitzada consisteix simplement a calcular la diferència entre dues mostres consecutives. En principi, des d'un punt de vista intuïtiu, aquesta transformació sembla natural a causa de la redundància que hi ha entre dos elements consecutius. Se suposa que, una vegada realitzada la transformació, la majoria de mostres siguin en un entorn pròxim al zero de manera que, segurament, els missatges amb valors pròxims a zero tindran moltes més probabilitats que els altres missatges.

Veiem que la transformació que hem utilitzat és invertible i que la informació original es pot recuperar de manera exacta a partir dels resultats transformats. Analíticament, la transformació diferencial queda representada per l'expressió

$$y[n] = x[n] - x[n-1],$$

mentre que la transformació inversa s'obté com a

$$z[n] = y[n] + z[n-1].$$

Es pot verificar que una transformació és inversa de l'altra només aplicant les equacions anteriors a les taules de valors que es mostren a la figura 17.

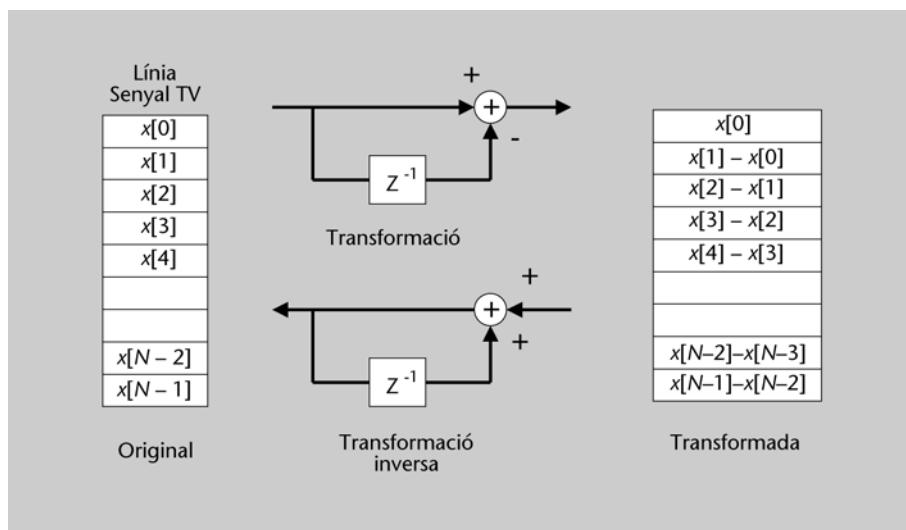


Figura 17. Exemple de transformació diferencial

Els resultats d'aplicar el codificador diferencial a una línia de la imatge es mostren a la figura 18, en què es representa el nivell de gris d'una línia original i els valors resultants després de la transformació. Podem veure com els valors transformats es concentren al voltant del zero i que, únicament en els contorns de la imatge, es produeixen valors significatius del senyal diferència.

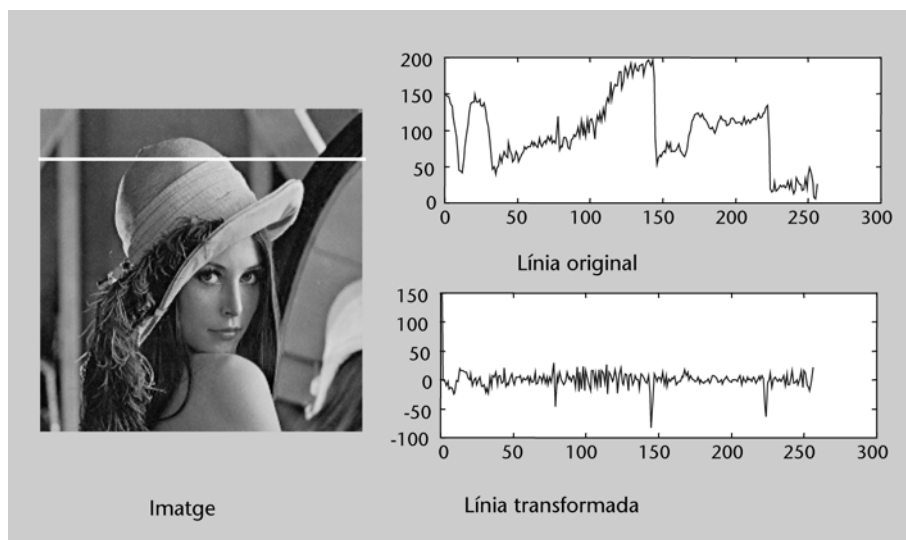


Figura 18. Representació dels nivells d'una línia original i una de transformada o diferencial

A la figura 19, es representen els resultats del codificador diferencial aplicats a una imatge en blanc i negre. La imatge original està codificada amb 256 nivells de gris (8 bits per píxel) i es representa assignant el negre al nivell 0 i el blanc al nivell 255. La imatge diferencial s'ha representat prenent el nivell 0 com un gris neutre. Els valors negatius són més foscos que el gris neutre, mentre que els valors positius són més clars. Aquesta imatge s'ha obtingut aplicant el codificador diferencial en cada una de les línies de la imatge. A partir d'aquests resultats visuals, és evident que hem aconseguit el propòsit que perseguíem,

és a dir: la majoria de píxels de la imatge diferencial són pròxims al nivell de gris mitjà i, només en els contorns de la imatge, apareixen píxels el valor dels quals s'aparta apreciablement del zero.



Figura 19. Imatge original i imatge diferencial

Aquest efecte es pot apreciar més clarament en les gràfiques de la figura 20 on es representen els histogrames de la imatge original i de la imatge diferencial. L'**histograma** és una representació bidimensional de la freqüència relativa amb què apareix cada nivell de gris en la imatge. Així, per a cada possible nivell de gris en l'eix d'abscisses, trobem el nombre de píxels de la imatge que prenen aquest nivell de gris. Per tant, l'histograma és una representació aproximada de la probabilitat que un píxel prengui un determinat nivell de gris en la imatge. Fixem-nos que, en aquestes gràfiques, s'han normalitzat els nivells de gris de manera que el negre correspon al zero mentre que el blanc correspon a la unitat. En l'histograma de l'esquerra, que es correspon amb la imatge original, els nivells de gris es distribueixen dins de tota la gamma sense mostrar una preferència directa per cap valor. En canvi, en la gràfica de la dreta, que correspon a la imatge diferencial, hi ha una gran concentració d'elements d'imatge amb un nivell de gris pròxim a 0,5 (gris neutre amb el qual hem representat el valor zero).

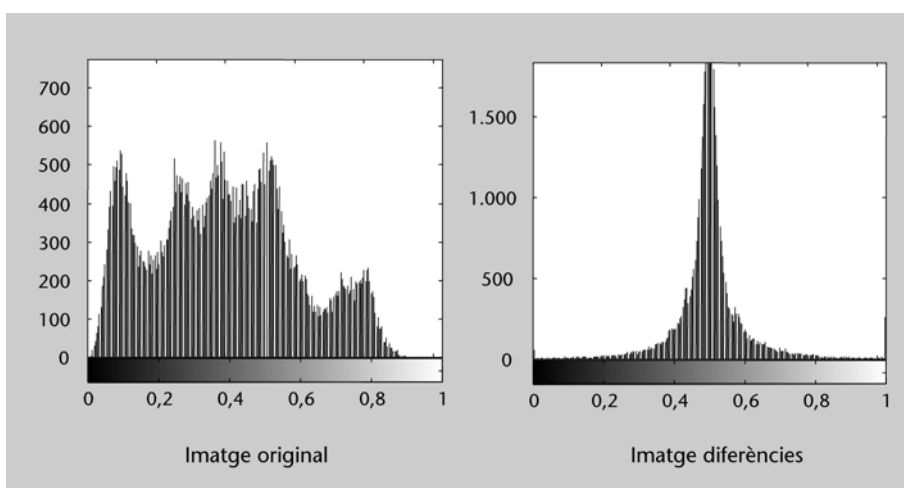


Figura 20. Histogrames de la imatge original i la imatge transformada o diferencial

Quan un senyal té una estadística en la qual predominen uns quants valors, que es presenten amb molta probabilitat, davant molts altres possibles nivells que es presenten amb poca probabilitat, és quan resulta rendible utilitzar codis de longitud variable en els quals els valors més probables es codifiquen emprant pocs bits, mentre que els valors amb poca probabilitat utilitzen una gran quantitat de bits. Per a tenir una idea del guany que hem obtingut en aplicar la transformació diferencial, podem comparar l'entropia de la font original i de la font diferencial, suposant que les probabilitats de cada missatge es poden aproximar pels histogrames calculats anteriorment. En aquest cas, obtenim que l'entropia de la imatge original és de 7,58 bits, mentre que l'entropia de la imatge diferencial és de 5,8 bits.

És important remarcar que la codificació que hem realitzat fins a aquest moment és una codificació sense pèrdues, sempre que mantinguem els càlculs de la diferència amb la precisió adequada. En aquest exemple, la imatge original estava codificada amb 256 nivells de gris (8 bits per mostra); per tant, els senyals diferència poden prendre valors compresos entre -255 i 255, ja que és possible que un píxel prengui el valor nul quan l'anterior era blanc o viceversa. Això significa que si volem reconstruir la imatge original amb els nivells de gris exactes (sense pèrdues) haurem de mantenir els 511 possibles valors que pot prendre el senyal diferència (9 bits per mostra). Si aquests valors s'aproximen mitjançant una quantificació més tosca, ja no serà possible recuperar la imatge original de manera exacta.

8.2. Predicció lineal. Una generalització de la codificació diferencial

La codificació diferencial es pot considerar un cas particular de la predicció de mostres. Els **predictors lineals** són uns dels mètodes més utilitzats per a la codificació de senyals d'àudio i vídeo, i el seu funcionament es basa en l'estimació del valor de la mostra actual a partir d'una combinació lineal de les mostres anteriors. La idea general és que combinant diverses de les mostres anteriors podem obtenir una estimació millor de la mostra actual. Per tant, si millorem l'estimació de la mostra, el senyal diferència tindrà una probabilitat molt alta per als missatges pròxims a zero i es podrà codificar de manera molt eficient mitjançant codis de longitud variable.

L'equació que regeix el filtre predictor

$$\tilde{x}[n] = \sum_{k=1}^P a_k \cdot x[n-k],$$

on P és l'ordre del predictor i a_k són els coeficients de predicció. Els coeficients de predicció poden determinar-se partint de les característiques estadístiques del senyal o ser establerts per endavant entre el codificador i el descodificador.

L'ordre del predictor...

... és el nombre de mostres anteriors que s'utilitzen per a realitzar la predicció de la mostra actual.

La idea bàsica per a codificar el senyal és que tant el transmissor com el receptor coneguin els coeficients que s'utilitzaran en el predictor i que tots dos utilitzin els mateixos coeficients a tota hora. Una possibilitat és que el transmissor calculi els coeficients i els envii al receptor. I una altra possibilitat és que tots dos calculin els coeficients, utilitzant el mateix algoritme i amb les mateixes mostres de senyal, per a tenir sempre la garantia que el predictor usat en el transmissor i l'usat en el receptor són el mateix. D'aquesta manera, tant el transmissor com el receptor utilitzaran aquests coeficients per a predir la mostra següent a partir de les mostres anteriors. Aquesta predicció té un error, que serà conegut pel transmissor i que l'enviarà al receptor perquè aquest pugui determinar el valor exacte de la mostra actual. Així, doncs, el senyal que es transmet és l'error de predicció que esperem que, amb molta probabilitat, prengui valors pròxims a zero.

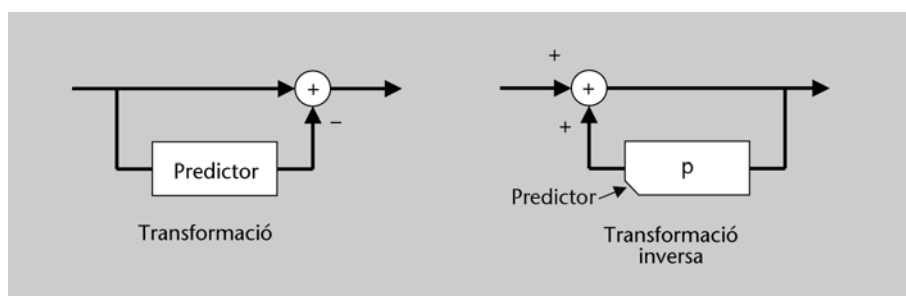


Figura 21. Esquema genèric d'un predictor i de la transformació inversa

La codificació diferencial...

... es pot considerar un cas especial d'un predictor que utilitza un únic coeficient (la mostra anterior) amb valor unitat.

8.3. Codificació diferencial i predicció d'un senyal d'àudio

Considerarem un exemple utilitzant un fragment de senyal d'àudio per a il·lustrar l'efecte de les transformacions de les dades sobre l'entropia. La figura 22 mostra la forma d'ona d'un fragment de senyal d'àudio d'aproximadament 3 segons de durada. El fragment té un sol canal (monofònic) i està mostrejat a 44.100 mostres per segon. La codificació original és en PCM utilitzant 8 bits per cada mostra.

A partir d'aquest fragment original realitzem la codificació diferencial, és a dir, la informació que transmetem o emmagatzemem és la diferència entre dues mostres consecutives. La forma d'ona del senyal diferència obtinguda es mostra a la figura 23. Hauríem de veure que les formes d'ona de tots dos senyals tenen una certa semblança que, d'altra banda, resulta lògica. De fet, l'energia del senyal diferència augmenta en els intervals en què l'energia del senyal original també augmenta, la qual cosa és perfectament natural, ja que el nostre error en l'estimació augmenta quan el nivell del senyal és major. A part d'això, apreciaríem diferències significatives en la forma d'ona precisa de tots dos senyals si els observéssim en detall (realitzant un *zoom* de les gràfiques per veure'n el comportament en un interval de curta durada). No obstant això, la diferència més important entre totes dues formes d'ona és el nivell de senyal. Una de les característiques que podem apreciar és que el senyal diferència té

un nivell molt inferior al del senyal original (escala de l'eix vertical), la qual cosa indica que la nostra estimació ha funcionat de la manera esperada.

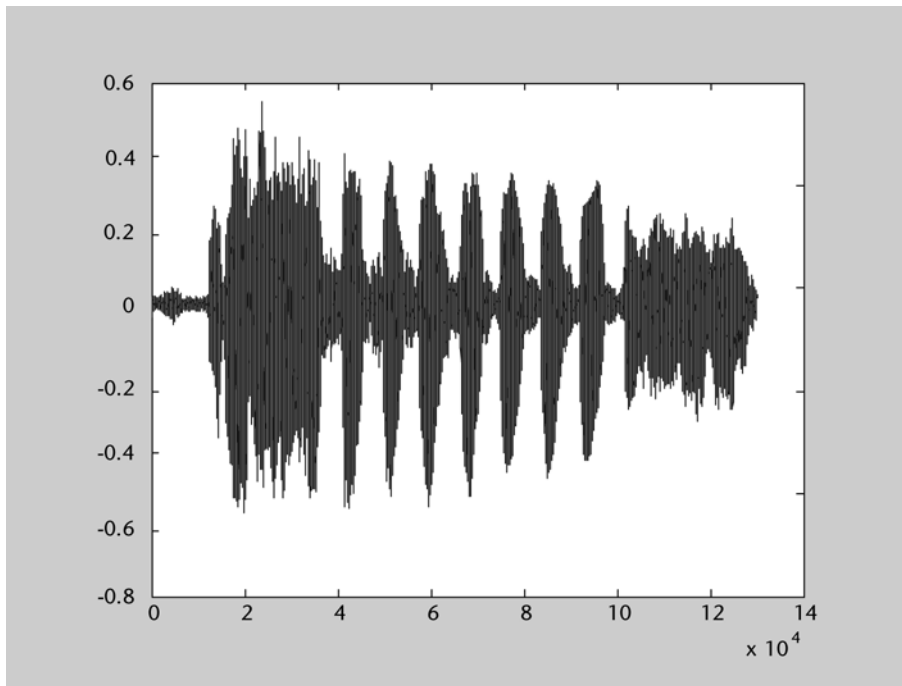


Figura 22. Forma d'ona d'un fragment d'àudio (fm = 44.100 Hz, 8 bits per mostra)

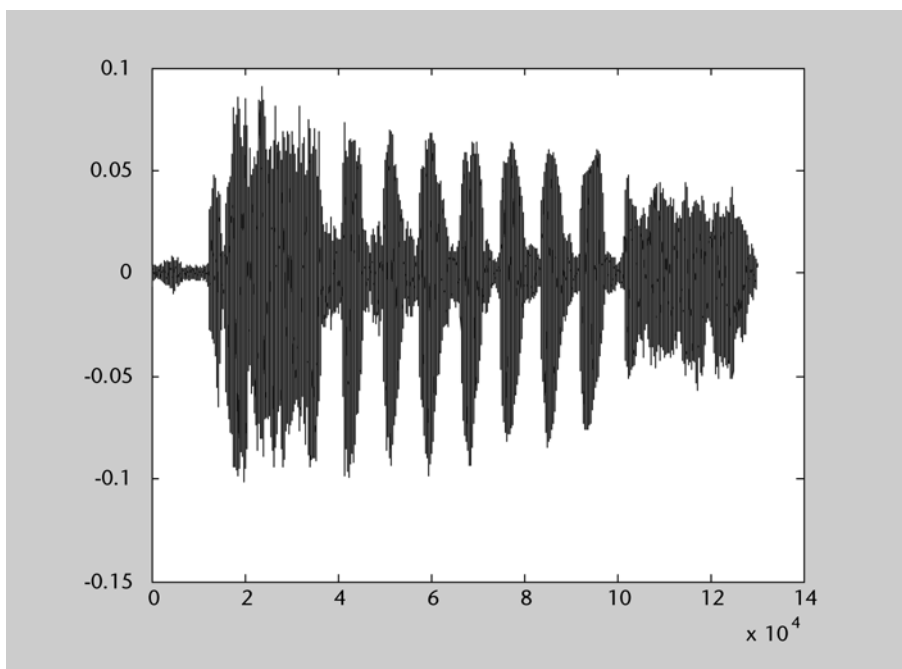


Figura 23. Senyal diferència obtingut a partir del fragment d'àudio original de la figura 20

Per a comparar aquests resultats, a la figura 24 es representen els histogrames de tots dos senyals resultants. Els histogrames s'han calculat suposant que els senyals resultants es divideixen en 256 possibles nivells de quantificació (256 missatges). En l'eix d'abscisses d'aquestes gràfiques apareix, per tant, el número del missatge, i no el valor numèric del senyal. A més, els histogrames se centren aproximadament sobre el missatge 128, que es correspon amb nivells de senyal pròxims a zero.

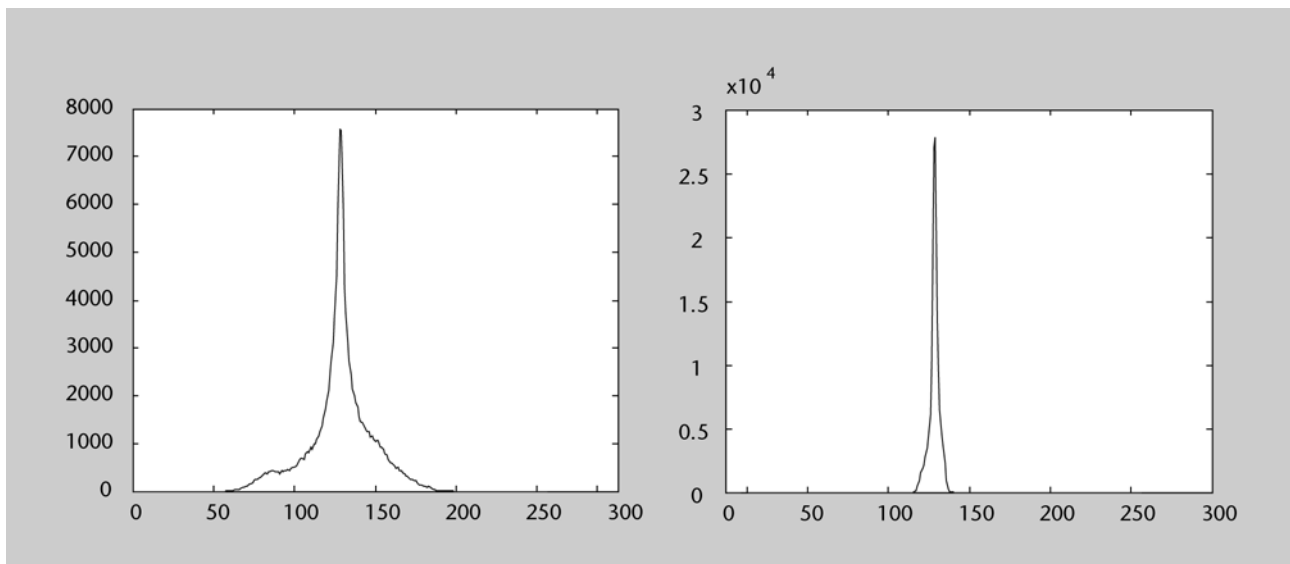


Figura 24. Comparativa entre l'histograma del senyal d'àudio original (esquerra) i el senyal diferència (dreta). L'histograma d'aquest últim té la majoria dels seus valors concentrats en la regió central.

Immediatament, es comprova que la transformació diferencial augmenta de manera considerable la probabilitat de transmetre els nivells més baixos. A més, si calculem l'entropia del senyal original i del senyal diferència obtenim 6,12 bits i 3,54 bits respectivament, la qual cosa ens indica que podrem trobar codis eficients per a codificar aquest últim senyal per sota de la taxa mitjana de bits necessària per a codificar el senyal original (els criteris de Huffman ens garanteixen que el senyal diferència es podrà codificar entre 3,54 bits i 4,54 bits per mostra de mitjana).

Per acabar, vegem el resultat que s'obté quan utilitzem un predictor. En aquest exemple concret, utilitzarem un predictor amb quatre coeficients, és a dir, que l'estimació de la mostra actual s'obté ponderant les quatre mostres anteriors amb uns coeficients $\{a_i\}$:

$$\tilde{x}[n] = \sum_{i=1}^4 a_i \cdot x[n-i]$$

En aquest cas, el senyal que transmetem o emmagatzemem serà la diferència entre la mostra original i la mostra estimada mitjançant l'equació anterior:

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

Els coeficients del predictor es poden calcular de diferents maneres. Una possibilitat, la més senzilla, consisteix a fixar-ne el valor per endavant d'acord amb raonaments intuïtius. Així, per exemple, podríem prendre tots els valors a_i iguals a 0,25, amb la qual cosa el valor estimat per a la mostra actual no és més que el valor mitjà de les mostres anteriors. Aquesta aproximació resulta bastant raonable quan sabem que el senyal varia molt lentament i que hi ha un nivell de soroll significatiu en les mostres. Una altra possible aproximació seria $a_1 = 0,4$, $a_2 = 0,3$, $a_3 = 0,2$, $a_4 = 0,1$, on també realitzem una mena de mitjana, però donant ara més pes a les mostres més recents.

Una altra manera d'enfocar el problema, més matemàtica i amb millors resultats, consisteix a determinar els coeficients del predictor que resultaran òptims per a aquest senyal concret. Per a resoldre aquest problema, haurem de definir el concepte d'*optimització* (*optimació*), és a dir, el criteri que utilitzem per a dir que un conjunt de coeficients determinat és millor que un altre. Generalment, el criteri d'optimització es pren com el valor mínim de l'error quadràtic mitjà del senyal diferència, és a dir:

$$\min \left\{ E[s^2[n]] = \frac{1}{M} \sum_n s^2[n] \right\}$$

Utilitzant aquest criteri de minimització, és possible determinar els coeficients del predictor òptim per al nostre fragment de senyal. En el nostre exemple, els coeficients que obtenim en resoldre aquest problema són els següents:

$$a_1 = 3,2528; a_2 = -4,1701; a_3 = 2,5477; a_4 = -0,6394$$

La figura 25 mostra esquemàticament com es construeix, a partir del senyal original, el senyal d'error (el senyal que transmetrem o emmagatzemarem).

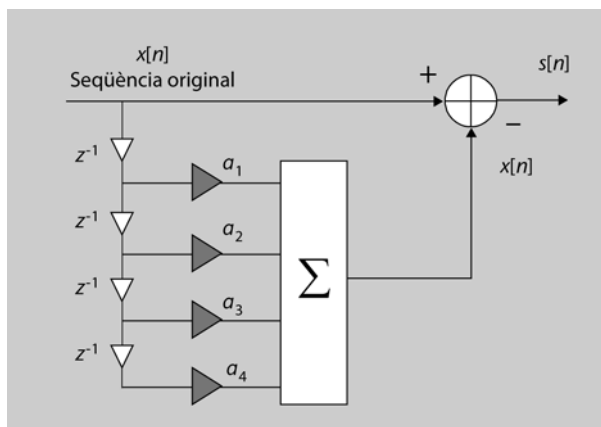


Figura 25. Obtenció del senyal diferència utilitzant un predictor lineal de quatre coeficients

A la figura 26, hem representat la forma d'ona del senyal $s[n]$ que s'obté en aplicar el predictor. Com en el cas de la codificació diferencial, la forma genèrica s'assembla a la del senyal original, ja que, com hem vist, totes dues han de tenir els màxims d'energia en els mateixos intervals. No obstant això, els detalls de la forma d'ona són diferents i, encara més important, el nivell de la seqüència $s[n]$ és considerablement menor que el del senyal original i que el del senyal diferència. Això ens indica que la predicció mitjançant quatre coeficients ha portat una millora respecte a la codificació diferencial.

La figura 27 mostra l'histograma del senyal $s[n]$, molt més concentrat en uns quants missatges que els histogrames que mostra la figura 24. Per a calcular aquest histograma, hem suposat que els valors del senyal es divideixen en 256 possibles nivells de quantificació.

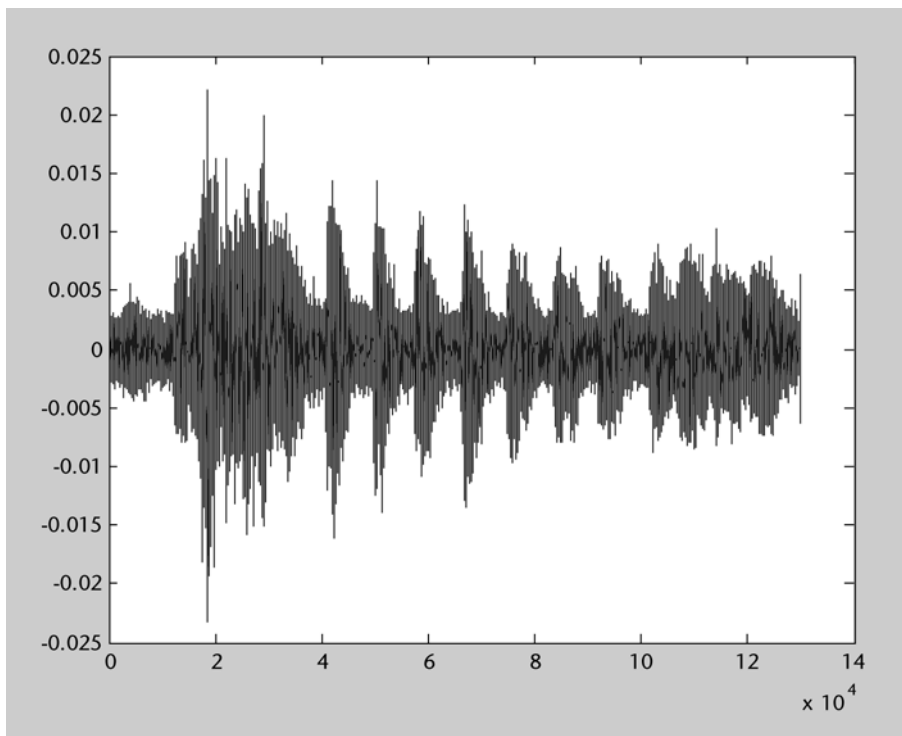


Figura 26. Forma d'ona del senyal $s[n]$ obtinguda amb l'ús d'un predictor de quatre coeficients

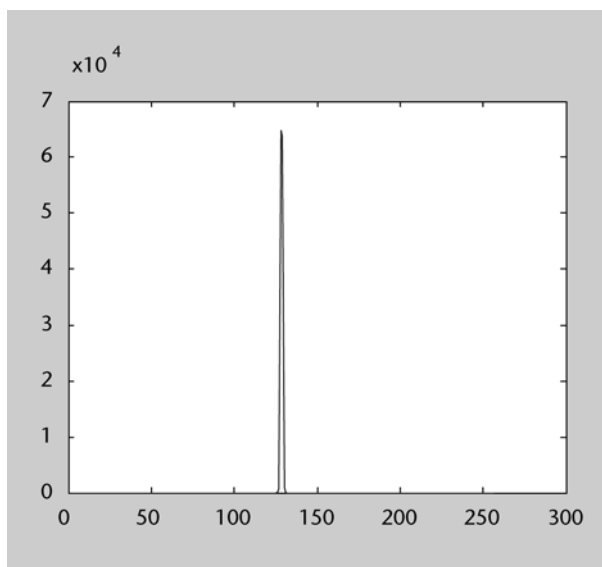


Figura 27. Histograma del senyal $s[n]$ obtingut com la diferència entre el senyal original i la predicció amb quatre coeficients

També s'inclou una taula comparativa de les entropies de les tres fonts: original, diferencial i diferencial amb predictor. Per determinar les entropies, hem estimat les probabilitats de les fonts a partir dels histogrames. Pot resultar curiós "escoltar" les formes d'ona del senyal original, el diferencial i la diferència de predicció utilitzats en aquest exemple*. En escoltar aquests senyals, es reconeix clarament la melodia original en els fitxers que contenen el senyal diferencial i l'error de predicció, encara que amb una qualitat considerablement inferior. Aquesta última té un nivell molt baix perquè la predicció és força bona i, per tant, el nivell de la diferència és molt reduït.

* Es proporcionaran els arxius RNoche_orig.wav, RNoche_dif.wav i RNoche_pred.wav en l'Aula de l'assignatura.

Codificador	Entropia
Senyal original (PCM)	6,12
Senyal diferència	3,54
Senyal diferència amb predicció 4 coeficients	1,09

Taula 3. Comparativa entre les entropies de totes tres fonts

Els codificadors que hem utilitzat en l'exemple anterior són codificadors amb pèrdues. El senyal original està codificat amb 8 bits en PCM (256 nivells) i, en conseqüència, el senyal diferència, és a dir, la resta entre dues mostres consecutives, pot prendre 511 valors diferents (qualsevol nombre enter entre -255 i 255). Per tant, quan tornem a aproximar els resultats obtinguts a 256 nivells, estem introduint pèrdues, amb la qual cosa el senyal original ja no es podrà recuperar de manera exacta.

En reduir el nombre de possibles missatges a transmetre, també es redueix l'entropia. Quan el nombre total de missatges es compacta a la meitat, l'entropia es redueix, com a màxim, en un bit. Així, per al nostre problema, es pot comprovar que l'entropia que hauríem obtingut si s'hagués mantingut el senyal diferencial sense errors (511 nivells) és de 4,51 bits. Aquest resultat ens indica que la codificació diferencial, encara que es realitzi sense pèrdues, aporta una millora en compressió respecte a la codificació PCM original, ja que aconseguim reduir l'entropia en aproximadament 1,5 bits.

8.4. Codificació diferencial adaptativa

En l'exemple de l'apartat anterior, el senyal d'àudio té una durada aproximada de tres segons, i hem utilitzat un únic predictor de quatre coeficients per a realitzar l'estimació de la mostra següent. Els coeficients del predictor s'han calculat perquè s'obtingui un valor quadràtic mínim del senyal diferència. Evidentment, aquesta idea es pot generalitzar i aplicar a fragments de senyal més petits, calculant el predictor òptim per a cada un d'aquests fragments. Aquesta filosofia de codificació s'utilitza molt i rep el nom de **codificació adaptativa** (ADPCM o *adaptive differential pulse code modulation*), ja que els coeficients del predictor es van modificant i adequant a les característiques del senyal al llarg del temps.

Vegem un exemple simplificat de com es podria estructurar un codificador amb aquestes característiques:

- 1) El codificador analitza el senyal en paquets de longitud preestablerta (per exemple 1.024 mostres).
- 2) Per a cada paquet, el codificador determina els coeficients del predictor òptim.
- 3) Es transmeten els coeficients del predictor òptim al descodificador. També es transmet el senyal diferència de predicció que s'ha obtingut en el codificador.
- 4) El descodificador recupera el senyal original a partir de les mostres anteriors, els coeficients del predictor i el senyal diferència.

Aquesta idea general admet múltiples variants tant en els paràmetres del sistema (longitud del paquet de mostres, nombre de coeficients del predictor, nombre de missatges amb què es quantifica el senyal diferència, etc.) com en l'estratègia de càlcul (admetre encavalcaments entre paquets, estratègies d'adaptació dels coeficients del predictor mostra a mostra, etc.).

9. Codificació de veu en telefonia mòbil GSM

En aquest apartat descriurem genèricament el sistema de codificació de canal que s'utilitza en telefonia mòbil GSM. Alguns detalls d'aquest sistema de codificació són molt complexos i sobrepassen els objectius d'aquest text, de manera que ens limitarem a exposar les idees més generals.

En aplicacions de telefonia, per a obtenir una qualitat prou satisfactòria mitjançant una codificació PCM, és necessari mostrejar el senyal de veu a raó de 8.000 mostres per segon prenent un total de 8 bits per mostra (256 nivells). Això significa que la taxa de bits per a telefonia digital PCM és la següent:

$$R = 8.000 \frac{\text{muestras}}{\text{s}} \times 8 \frac{\text{bits}}{\text{muestra}} = 64.000 \text{ bps} .$$

L'XDSI (xarxa digital de serveis integrats) i altres sistemes de multiplexació troncal de canals digitals utilitzen aquest tipus de codificació per a transmetre el senyal de veu. No obstant això, en telefonia mòbil digital, aquest flux de dades és elevat i és convenient comprimir-lo.

A més, hem de tenir en compte que el sistema de compressió no pot ser excessivament complex, ja que els processadors dels telèfons mòbils han de ser simples i de cost reduït: una complexitat computacional major es tradueix en un consum d'energia major, i la durada de la bateria és una de les característiques més apreciades en un telèfon mòbil.

Quan es va elaborar l'estàndard GSM es van estudiar i avaluar diferents alternatives per a la codificació del senyal de veu. Principalment, es van tenir en compte la qualitat subjectiva i la complexitat. Finalment, es va escollir un codificador basat en un vocodificador o *vocoder* de predicció lineal, l'RPE-LPC-LTP (*regular pulse excited-linear predictive coder with long term prediction*), al qual es van afegir algunes millores addicionals. La idea general d'aquest codificador consisteix a dividir el senyal de veu en fragments de curta durada: cent seixanta mostres, el que correspon a 20 ms. Cada un d'aquests paquets es codifica amb 260 bits, la qual cosa permet codificar el senyal d'àudio a una taxa de 13.000 bps.

Per a cada paquet:

- 1) Es determinen els coeficients del predictor òptim juntament amb una possible forma d'ona que permeti reproduir el senyal d'entrada en el decodificador.

2) Es codifiquen els coeficients del predictor més els paràmetres de la forma d'ona amb què s'excitarà el predictor.

Finalment, la part *long term prediction* del codificador es refereix a la utilització d'una predicció addicional de la forma d'ona a partir dels paquets que s'han transmès.

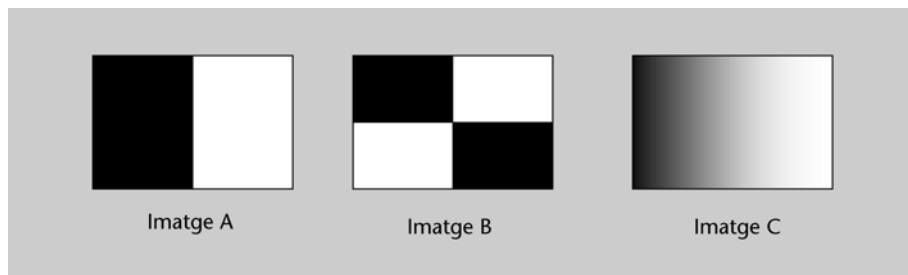
Activitats

1. Intenteu justificar per què els senyals diferència i diferència de predicció esmentats al final de l'apartat 8.3 són intel·ligibles.

2. Una font pot proporcionar vuit missatges amb les probabilitats següents: 0,6; 0,1; 0,1; 0,1; 0,025; 0,025; 0,025; 0,025. Es demana el següent:

- Determineu el codi de Huffman.
- Determineu l'entropia de la font.
- Determineu el nombre mitjà de bits que utilitzarà el codi de Huffman.

3. Considereu les tres imatges de la figura 28 i indiqueu si les afirmacions següents són veritades o falses.



La imatge C és la que té més entropia.	
La imatge A i la B tenen la mateixa entropia.	
La imatge B és la que té més entropia.	
L'entropia de la imatge A és d'1 bit.	

4. Comproveu que, per a la font de la figura 12, l'entropia és 2,344 i que el nombre mitjà de bits del codi que s'obté a la figura 11 és de 2,4.

5. Considereu el codi de longitud variable que es proporciona en aquesta taula:

Missatge	Codi	Missatge	Codi
M_1	00	M_5	101
M_2	010	M_6	110
M_3	011	M_7	1110
M_4	100	M_8	1111

a) Determineu la paraula codi amb què es codificaria la seqüència de missatges següent:

$$M_3 M_5 M_1 M_8 M_1 M_1 M_6 M_7$$

b) Determineu la seqüència de missatges a què correspon la seqüència de bits següent:

$$010100101111000000101111100000$$

c) Determineu el valor mitjà de bits del codi suposant que tots els missatges són equiprobables. Raoneu intuïtivament per què el resultat és superior a 3.

6. Considereu el codi de longitud variable proporcionat en la taula de l'exercici anterior. Determineu quina condició han de complir les probabilitats d' M_1 , M_7 i M_8 perquè el codi resultant tingui una taxa de bits mitjana menor que 3.

7. Determineu quant temps es necessitaria per a transmetre una imatge de 3.000 files per 4.000 columnes, codificada en PCM amb 24 bpp (bit per píxel), per una línia ADSL d'1 Mbps. Suposeu que el rendiment de la línia ADSL és màxim i que s'aprofita tot el flux d'1 Mbps per a transmetre la informació útil.

Bibliografia

Bàsica

Benedetto, S.; Biglieri, E. (1999). *Principles of Digital Transmission*. Kluwer Academic Press / Plenum publishers.

Proakis, J. G. (2003). *Digital Communications* (4a. ed.). McGraw Hill.

Complementària

Carlson, A. B. (2001). *Communication Systems: An Introduction to Signals and Noise in Electrical Communication* (4a. ed.). McGraw Hill.

Stix, G. (1991, setembre). "Encoding the Neatness of Ones and Zeroes". *Scientific American* (pàg. 54-58).

Segal, J. "El geòmetra de la informació". *Temas Investigación y Ciencia: La Información* (núm. 36, pàg. 12-15).

Bekenstein, Jacob D. "La información en el universo holográfico". *Temas Investigación y Ciencia: La Información* (núm. 36, pàg. 16-23).

Gibson, Jerry D. i altres (1998). *Digital Compression for Multimedia: Principles & Standards*. Morgan Kauffman.

Tarrés, F. (2001). *Sistemas Audiovisuales I: Televisión Analógica y Digital*. Ediciones UPC.