

CONT 2/5 : Contenus multimédia

Olivier Fourmaux

Version 1.4

Nombreux transparents/images empruntés à :

- Jean-Loup Guillaume
- Timur Friedman
- Anne Fladenmuller
- Wikipedia...

CONT : plan du cours 2/5

- 1 Introduction
 - contenus numériques
 - encodage des textes
 - encodage des images
- 2 Contenus audio
 - audio numérique
 - encodages sans pertes
 - compressions avec pertes
- 3 Contenu vidéo
 - vidéo numérique
 - encodages sans pertes
 - compressions avec pertes

CONT : plan du cours 2/5

- 1 Introduction
 - contenus numériques
 - encodage des textes
 - encodage des images
- 2 Contenus audio
 - audio numérique
 - encodages sans pertes
 - compressions avec pertes
- 3 Contenu vidéo
 - vidéo numérique
 - encodages sans pertes
 - compressions avec pertes

Compression



Definition

Transformer les données associées au contenu en une représentation plus compacte

Pourquoi compresser ?

- réduction de l'espace de stockage
- réduction du temps de transmission (texte, images, VoD...)
- limitation du débit (audio, vidéo...)

Compression : exemple simpliste

Texte codé en ASCII (1 caractère codé sur 1 octet) :

- fichier texte comportant 100 caractères 'a'
 - taille du fichier = 100 octets
- stockage sous la forme de la valeur 100 codée sur un octet suivi du caractère 'a'
 - taille du fichier = 2 octets
 - même information
 - 2/100 \Rightarrow taux de compression = 2%, taux de réduction = 98%

Compression : terminologie

Compression sans perte (codage réversible)

- ➔ les données originales peuvent être reconstituées
 - intégrité des données originales
 - ex : compress, zip, gzip...

Compression avec perte (codage non réversible)

- ➔ les données originales sont perdues
 - limitation de l'impact selon la perception humaine
 - taux de compression variable
 - ex : formats JPEG, MP3, H264...

Compression spatiale, temporelle...

- spécifique aux images, à l'audio ou à la vidéo

Différents types de contenus numériques

Dans la suite nous allons présenter :

- Contenus statiques (indépendants des aspects temporels)
 - texte
 - image
- Contenus dynamiques (lecture associée à un interval de temps, contrainte temporelle pour leur transport)
 - audio
 - vidéo

CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

3 Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes

Historique



- Premiers codages de caractères
 - 1794 : Signaux télégraphique Chappe (visuels)
 - 1832 : Alphabet Morse (télégraphe électrique)
 - 1928 : Carte perforée IBM 80 colonnes (12 lignes)
- Codages de caractères par bits
 - 1874 : Baudot 5 bits (premiers téléscripteurs)
 - 1954 : BCD 6 bits (adaptation binaire des cartes IBM)
 - 1963 : ASCII 7 bits et EBCDIC 8 bits
 - 198x : extensions multilingues sur 8 bits
 - 199x : Unicode (16 bits, 32 bits...)

Contenu textuel

Informations textuelles utilisent des ensembles de caractères abstraits

- information déjà discrétisée
- un caractère = un son, une syllable ou un concept
 - alphabets, syllabaires ou idéogrammes
- ensemble de caractères = jeux de caractères (*charset*)

Codage des caractères = représentation par un mot de code (numérique) de chaque caractère d'un jeu de caractères

- nombreuses représentations pour chaque jeu de caractère : il faut toujours préciser le l'encodage utilisé
- pour un texte, toujours préciser :
 - le jeu de caractère (US-ASCII, fr, latin1, cyrillic, jp...)
 - l'encodage (EBCDIC, ISO-646, ISO-8859-1, UTF-8...)

ASCII

bits	0	0	0	1	0	1	0	1	1	1
bits	0	0	1	1	0	1	0	1	0	1
bits	CONTROL		SYMBOLS NUMBERS		UPPER CASE		LOWER CASE			
0 0 0 0	NUL	DLE	SP	0	@	P		p		
0 0 0 1	SOH	DC1	!	1	A	Q	a	q		
0 0 1 0	STX	DC2	"	2	B	R	b	r		
0 0 1 1	ETX	DC3	#	3	C	S	c	s		
0 1 0 0	EOT	DC4	\$	4	D	T	d	t		
0 1 0 1	ENQ	NAK	%	5	E	U	e	u		
0 1 1 0	ACK	SYN	&	6	F	V	f	v		
0 1 1 1	BEL	ETB	'	7	G	W	g	w		
1 0 0 0	BS	CAN	(8	H	X	h	x		
1 0 0 1	HT	EM)	9	I	Y	i	y		
1 0 1 0	LF	SUB	*	:	J	Z	j	z		
1 0 1 1	VT	ESC	+	;	K	[k	{		
1 1 0 0	FF	FS	=	<	L	\	l			
1 1 0 1	CR	GS	-	=	M]	m	}		
1 1 1 0	SO	RS	.	>	N	^	n	~		
1 1 1 1	SI	US	/	?	O	_	o	DEL		

American Standard Code for Information Interchange

- ANSI X3.4 (standard pour l'administration US)
- IANA : US-ASCII
- 7 bits
- 33 caractères de contrôle
- 95 caractères imprimables
- alphabet anglais

Extensions internationales de l'ASCII

Nombreuses variantes sur 8 bits pour un nombre limité d'alphabet :

- encodages DOS : CP437 (latin), CP667 (polish), CP720 (arabic), CP850 (europe occ.), CP851 (greek), CP852 (central europe), CP866 (cyrilic), CP936 (simplified chinese)...
- encodages MacOS : Roman, CentralEurRoman, Greek, Cyrillic, Japanese, Korean, ChineseSimp, ChineseTrad, Thai/TIS-620...
- encodages Windows : CP1250, CP1251, CP1252, CP1253... CP874/TIS-620, CP932/ShiftJIS, CP936/GBK, CP949/EUC-KR, CP950/Big5...
- normes nationales : GOST, ISCII, TSCII, VISCII, YUSCII, PASCII, HKSCS...

ISO/CEI 8859

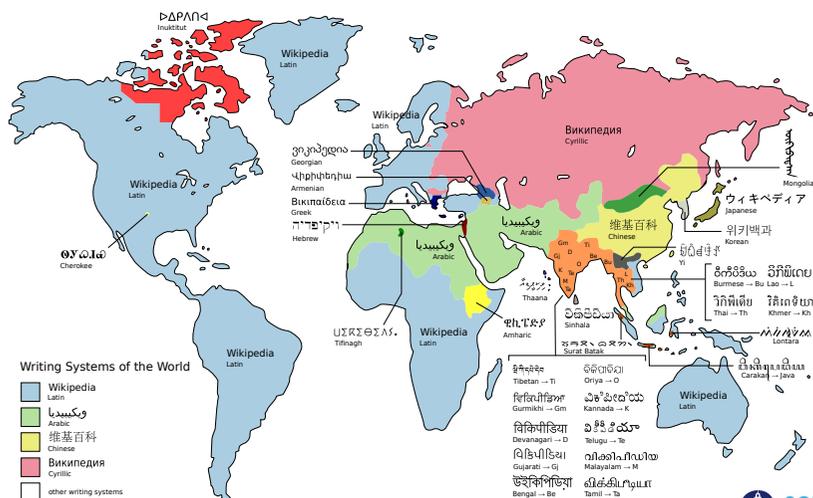
Norme internationale pour limiter les encodages (1985) :

- encodage sur 8 bits (compatible ASCII, pratique à manipuler)
- principe des pages de code DOS/windows
- originellement 6 codages latins + cyrillique, arabe, grec et hébreu :
 - ISO 8859-1 (iso-latin1) europe occidentale
 - ISO 8859-2 (iso-latin2) europe centrale
 - ...
- Le français est complètement géré dans une extension (1998) :
 - ISO 8859-15 (iso-latin9) = iso-latin-1 avec €, Œ, œ et Ý



mais de nombreuses écritures ne sont pas prises en compte...

Différentes écritures du monde



ISO/CEI 10646

Tentative de standardisation d'un jeu de caractère unique :

- approche **globale** : Jeu universel de caractères codés (JUC/UCS)
- identification de plus de **136755 caractères** prenant en charge 139 écritures (en 2017...)
- identifiant **hexadécimal** unique (codepoint compris entre U+0000 et U+10FFFF, 21 bits, certaines valeurs réservées)
 - "ç" : U+00E7
 - "ç" : LETTRE MINUSCULE LATINE C CÉDILLE
LATIN SMALL LETTER C WITH CEDILLA
- encodages envisagés sur 2 octets (UCS-2, plan de base) ou 4 octets (UCS-4)

Jeu universel de caractères codés (JUC/UCS)

Basic Latin ASCII	Thai	Linear B Syllab.	General Punctuation
Latin-1 Suppl.	Lao	Linear B Ideogr.	Currency Symbols
Latin Extensions	Tibetan	Lycian	Arrows
IPA Extensions	Burmese	Carian	Mathematical Operators
Greek and Coptic	Georgian	Old Italic	Enclosed Alphanumerics
Cyrillic	Hangul jamo	Gothic	Box Drawing
Armenian	Ethiopic	Ugaritic	Block Elements
Hebrew	Cherokee	Old Persian	Geometric Shapes
Arabic	Can. Aboriginal	Deseret	Dingbats
Syriac	...	Shavian	Braille Patterns
Thaana	CJK Unif. Ideogr.	Osmanya	Musical Symbols
N'Ko	CJK Radicals	Cypriot Syllab.	Emoji
Samaritan	CJK Symbols	Imperial Aramaic	Alchemical Symbols
Mandaic	CJK Punctuation	Phoenician	Game Symbol
Devanagari	Kangxi Radicals	Lydian	
Bengali	Hiragana	...	
Gurmukhi	Katakana	Cuneiform	
Gujarati	Bopomofo	Egypt.	
...	Yi Syllables	Hieroglyphs	

Unicode

Consortium industriel (initié par Apple et Xerox, fin 80)

- initialement : écritures modernes seulement
 - Unicode 1.0 (1 plan, codage 16 bits, incompatible ISO 10646)
- convergence Unicode et ISO 10646 dès Unicode 1.1
 - utilisation du **même JUC/UCS**
- dernière version : Unicode 10.0 (juin 2017)
 - **136755 caractères** prenant en charge 139 écritures
- principaux encodages Unicode :
 - UTF-16 (mots de 16 bits, taille variable, pb d'ordre des octets)
 - UTF-32 (mots de 32 bits, taille fixe, pb d'ordre des octets)
 - **UTF-8** (taille variable, basé sur des octets)...

UTF-8

Format d'encodage de tous les codepoints Unicode/ISO 10646

- encodage par défaut des systèmes Unix et windows modernes
- ≈ 80% du codage du contenu textuel du web
- obligatoire pour tout nouveau protocole standard IETF

Codage sur 1 à 4 octets (reste compatible avec ASCII) :

codepoints	représentation binaire	premier octet
U+0000 à U+007F	0xxxxxxx	00 à 7F
U+0080 à U+07FF	110xxxxx 10xxxxxx	C2 à DF
U+0800 à U+FFFF	1110xxxx 10xxxxxx 10xxxxxx	E0 à EF
U+010000 à U+1FFFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx	F0 à F4

certaines séquences sont invalides

Ex. :	car.	codept.	déc.	binaire (octets)	hexadéc.
	A	U+0041	65	01000001	41
	é	U+00E9	233	11000011 10101001	C3 A9
	€	U+20AC	8364	11100010 10000010 10101100	E2 82 AC

Codage compressif du texte

Impératif de compression sans perte (intégrité des données)

Terminologie :

- **encodage** : chaque caractère à coder est représenté par un mot de code
- **mot de code** : composé de signes élémentaires
- **valence** : nombre de signes élémentaire utilisé
 - un code est binaire si il est de valence 2

Codes de longueur fixe ou variable...

Codes de longueur fixe

Tous les mots de code sont de longueur fixe

- ASCII, iso-latin-1, codepoint Unicode, UTF-32...

S'il y a n caractères à coder avec une valence v :

- nombre minimum de signe élémentaires : $\lceil \log_v(n) \rceil$
- exemple avec 6 caractères $\{a, b, c, d, e, f\}$ à coder en binaire :

- binaire $\implies v = 2$

- $\lceil \log_2(6) \rceil = \lceil 2,58... \rceil = 3$ bits par caractères

- un codage possible :

caractère	a	b	c	d	e	f
long. fixe	000	001	010	011	100	101



codes réservées (ex : codes de contrôles en ASCII)

Codage compressif sans perte

Objectifs :

- **codage intelligible** : possibilité de le décoder de manière unique
- **codage efficace** : les caractères les plus fréquents sont codés par des mots de code les plus court

Limites (codes binaires) :

- **Entropie** $H = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i}$
- **Théorème du codage de la source**
 - \exists un code avec une longueur moyenne à moins d'1 bit de H
 - \exists des codes aussi proche que l'on veut du minimum théorique
 - groupement de caractère nécessaire
 - l'efficacité nécessite de prendre en compte la fréquence...

Codes de longueur variable

Les mots de code ont des tailles variables selon les caractères

- un code est :
 - **intelligible** s'il a un unique décodage
 - **préfixe** si aucun mot de code n'est préfixe d'un autre
 - **instantané** si on peut le décoder au fur et à mesure

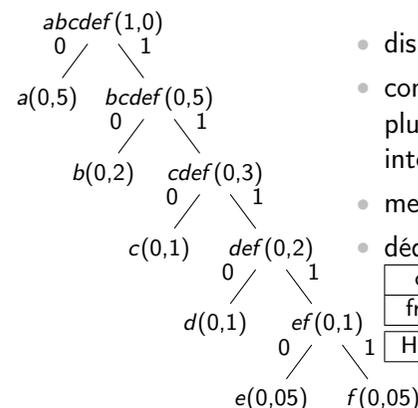
caractère	a	b	c	d	e	f
fréquence	0,5	0,2	0,1	0,1	0,05	0,05

- exemple :

long. fixe	000	001	010	011	100	101
long. var. 1	01	10	100	101	1011	0011
long. var. 2	0	101	100	111	1101	1100

- **long. var. 1** : 10101 code bd ou $da \implies$ non intelligible
- **long. var. 2** : préfixe \implies intelligible

Codage de Huffman



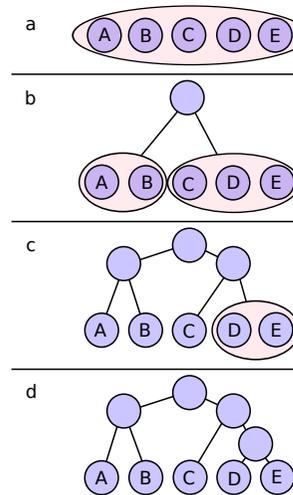
- disposer les caractères par fréquence
- connecter récursivement les deux nœuds de plus petite probabilité en créant les nœuds intermédiaires
- mettre un label sur les branches
- déduire le codage :

car.	a	b	c	d	e	f
fréq.	0,5	0,2	0,1	0,1	0,05	0,05
Huff.	0	10	110	1110	11110	11111

Caractéristiques du codage de Huffman

- le code utilise un nombre minimal de bits par caractères :
 - optimal si les fréquences sont des puissances inverses de 2 : 1/2, 1/4, 1/8, 1/16...
 - peut être éloigné de la borne théorique
 - amélioration en utilisant des **mots de codes**
 - exemple avec $\{a, b\}$ avec $p(a) = 0,6$ et $p(b) = 0,4$
 - $a = 0$ et $b = 1$: longueur moyenne = $0,6 * 1 + 0,4 * 1 = 1$
 - $aa = 00$, $ab = 01$, $ba = 10$ et $bb = 11$: longueur moyenne = 1
 - $aaa = 000$, $aab = 110$, $aba = 100$, $abb = 101$, $baa = 010$, $bab = 011$, $bba = 1110$ et $bbb = 1111$: longueur moyenne = 0,981
 - limite théorique = 0,97
- le **dictionnaire** doit être envoyé avec le code
- peut-être amélioré dynamiquement si les fréquences sont variables dans les données \Rightarrow **codage adaptatif**

Codage de Shannon-Fano



- associer les fréquences d'apparition aux caractères
- tri puis séparer récursivement en deux ensembles de tailles aussi proches que possible
- mettre un label sur les branches
- déduire le codage :

caractère	a	b	c	d	e
fréquence	0,38	0,18	0,15	0,15	0,13
Sh.-Fano	00	01	10	110	111

Codage par dictionnaire

Typiquement LZ ou LZW (Lempel-Ziv-Welch, 1984) :

- algorithme de compression de données sans perte
- dictionnaire** de caractères créé dynamiquement
- initialement le dictionnaire contient les 256 caractères ASCII
- extension dynamique** du dictionnaire avec des mots de code
- fichier codé usuellement avec des valeurs sur 12 bits
- le dictionnaire est reconstruit à la décompression
- exemple d'utilisation : Unix compress, format GIF

Compression LZW

in	add dict.	out	code (12b)
"i"			
"t"	"it"(256)	"i"	105
"t"	"tt"(257)	"t"	116
"y"	"ty"(258)	"t"	116
" "	"y "(259)	"y"	121
"b"	" b"(260)	" "	32
"i"	"bi"(261)	"b"	98
"t"			
"t"	"itt"(262)	"it"	256
"y"			
" "	"ty "(263)	"ty"	258
"b"			
"i"	" bi"(264)	" b"	260
"t"			
" "	"it "(265)	"it"	256

```

w = Nul
tant_que (lecture car. c) faire
  si (w + c existe dans le dict.)
    w = w + c;
  sinon
    ajouter w + c au dict.
    écrire le code de w
    w = c
  fin si
fin tant_que
écrire le code de w
  
```

Conclusion sur la compression de texte

Approches simples pour compresser :

- Huffman (gzip, JPEG)
- Shannon-Fano
- LZ/LZW (compress, zip, GIF)

Toutes reposent sur des algorithmes sans pertes

- ne nécessitent aucun a priori sur les données
- compression peu efficace
- adaptation aux contenus image/audio/vidéo ???

Images numériques

Les images sont des contenus analogiques liés à la perception visuelle humaine

Leur numérisation tiens compte de nombreux paramètres liés à celle-ci :

- définition spatiale
- intensités lumineuses
- valeurs chromatiques...

CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

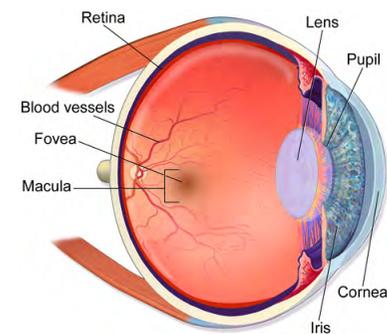
2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

3 Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes

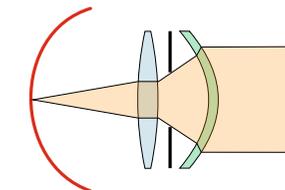
L'oeil humain



(CC BY 3.0 Bruce Blaus)

Système d'acquisition optique avec :

- lentilles (cornée et cristalin)
- diaphragme (iris)
- capteur (rétine)



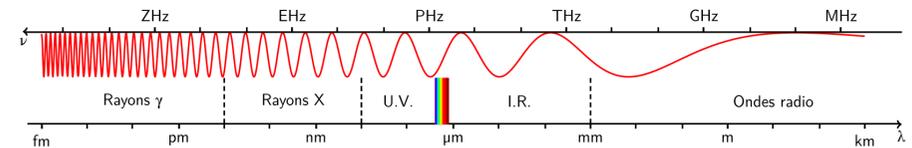
Récepteurs photo-sensibles de la rétine

Deux catégories de récepteurs décomposent les informations lumineuses :

- 120 millions de **bâtonnets**
 - vision crépusculaire et nocturne en noir et blanc
 - perception des formes et mouvements
- 6 millions de **cônes**
 - vision diurne
 - perception des couleurs
 - 3 types de cônes disposés en mosaïque aléatoire :
 - **B** (bleu) pic de sensibilité à la long. d'onde ≈ 420 nm (seul. 5%)
 - **V** (vert) pic de sensibilité à la long. d'onde ≈ 534 nm
 - **R** (rouge) pic de sensibilité à la long. d'onde ≈ 564 nm

Spectre électromagnétique

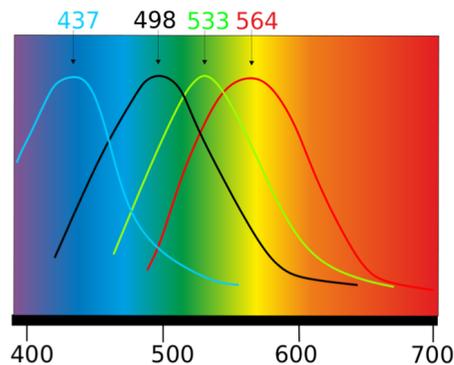
Régions approximatives en fréquence et en longueur d'onde du spectre électromagnétique :



(CC BY-SA 3.0 Benjamin Abel)

Spectre visible

Spectre d'absorption des trois types de cônes et des bâtonnets, récepteurs photo-sensible de la rétine humaine :



(CC BY-SA 3.0 Pancrat)

Paramètres de la couleur

Luminance(intensité lumineuse)

- Energie rayonnant d'une source jusqu'à l'œil (amplitude de l'onde)
- Luminosité (ou clarté) relative perçue
 - clair ou sombre
 - noir = absence de lumière reçue
 - blanc = toutes les lg d'onde au max.

Teinte (champ chromatique)

- longueur d'onde dominante
 - type de "couleur" (bleu, jaune...)

Saturation

- fonction inverse de la largeur de la bande du spectre
 - pâle ou vive



Photo-sensibilité de l'Œil

300.000 nuances de couleur distinguables (CIE)

Luminance

- sensibilité élevée

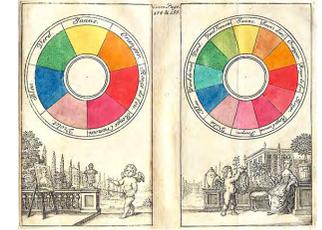
Chrominance (variation de couleur orthogonale à la luminance : combinaison de la teinte et de la saturation)

- sensibilité faible
- 128 teintes
 - sensibilité maximum pour le jaune
- 20 saturations
 - sensibilité maximum pour le rouge violacé

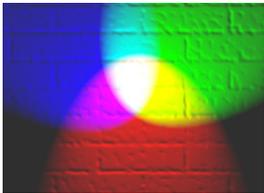
Perception visuelle des couleurs

Perception des couleurs par le cerveau humain :

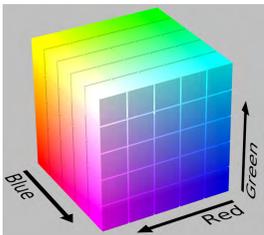
- pas que des couleurs pures
- des mélanges de fréquence atteignent les yeux
- ces mélanges sont eux-même perçus comme des couleurs
- la correspondance entre mélange et perception est complexe
 - plusieurs modèles explicatifs sont proposés : RGB, TSL, TSB, CIE RGB, CIE XYZ, CIE xyY , CIE Luv, CIE Lab, CMJ/CMJK, YIQ, YUV, YDbDr...



Système colorimétrique physiologique : RVB



(CC BY-SA 3.0 en :User :Bb3cxv)



(CC BY-SA 3.0 Pancrat)

Correspondance avec les 3 longueur d'onde des cônes de l'oeil humain

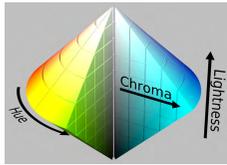
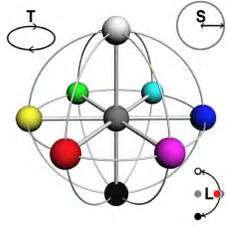
- synthèse additive
- complexe à manipuler (choix des couleurs)
- représentation partielle des couleurs perceptibles

Synthèse additive



Addition de la luminance de chaque composante

Système colorimétrique perceptif (1) : TSL/TLS



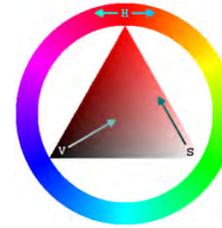
(CC BY-SA 3.0 SharkD)

Teinte, Saturation, Luminosité

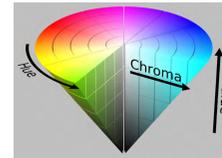
- modèle proposé au SIGRAPH'78 (HSL/HLS en anglais)
- paramètres :
 - Teinte = cercle des couleurs avec $0^\circ = 360^\circ = \text{rouge}$, $120^\circ = \text{vert} \dots$
 - Saturation = distance au gris (0%) à la teinte pure (100%)
 - Luminosité = position entre le noir (0%) et le blanc (100%)
- utilisé par la norme CSS et les logiciels Microsoft, CorelDRAW, AutoCAD, Inkscape...



Système colorimétrique perceptif (2) : TSB/TSV



(CC BY-SA 3.0 Samus)



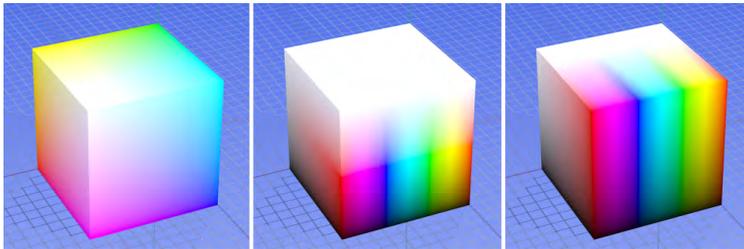
(CC BY-SA 3.0 SharkD)

Teinte, Saturation, Brillance/Valeur

- modèle également proposé au SIGRAPH'78 (HSB/HLV)
- paramètres :
 - Teinte = cercle des couleurs identique à TSL
 - Saturation = également distance au gris mais calcul différent à TSB
 - Brillance/Valeur = brillance de 0%(noir) à 100%
- utilisé par les logiciels Adobe, Apple, LibreOffice, Gimp...



Comparaison des systèmes RGB, TSL et TSV

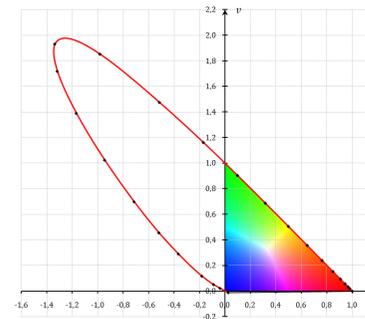


(CC BY-SA 3.0 SharkD)

- différents systèmes = différents espaces colorimétriques
- transformation non linéaire entre RVB et TSL/TSV



Normalisation CIE



(CC BY-SA 3.0 Alasjourn)

Modèle CIE RGB défini par la **commission internationale de l'éclairage (CIE)** en 1931

Couleur	Rouge R	Vert G	Bleu B
Lg d'onde	700,0 nm	546,1 nm	435,8 nm
Lumin. rel.	1,0000	4,5907	0,0601

Toutes les couleurs ne sont pas représentées dans le diagramme de chromaticité (courbe rouge = couleurs monochromatiques)



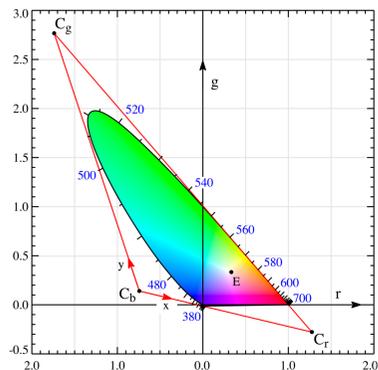
Système colorimétrique CIE XYZ

Nouveau système de coordonnées

- X, Y et Z (nouvelles "couleurs" primaires)
- correspondent à la réponse des différents cônes
 - chaque cône couvre un spectre recouvrant de longueurs d'ondes
 - Y est associé à la luminance
- produit toutes les couleurs perceptibles

Traduction entre RGB et XYZ :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.77 & 1.75 & 1.13 \\ 1.00 & 4.59 & 0.06 \\ 0.00 & 0.06 & 5.60 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$

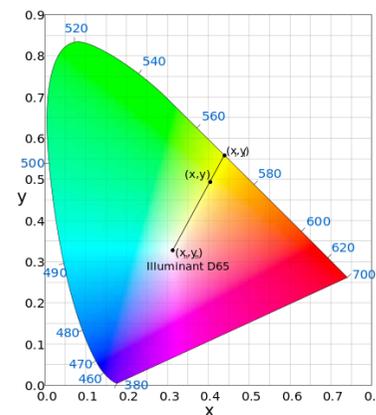


(CC BY-SA 3.0 PAR)

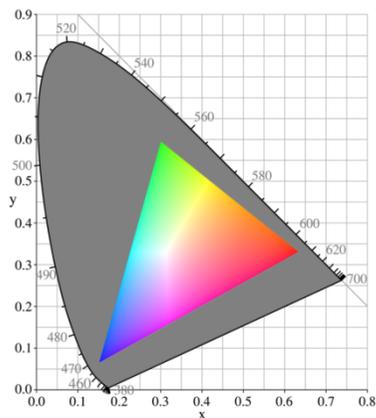
Système colorimétrique CIE Yxy

Dissociation de la luminance et de la chrominance

- Y = Luminance du système XYZ
- x et y caractérisent la couleur indépendamment de son intensité (avec $x = \frac{X}{X+Y+Z}$ et $y = \frac{Y}{X+Y+Z}$)
- permet de représenter les couleurs dans un diagramme de chromaticité xy



Gamut d'un système d'affichage

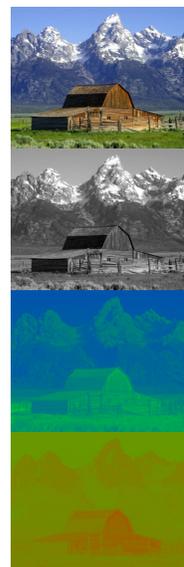


Gamut = étendue des couleurs affichables

- intérieur du **triangle** pour un système d'affichage en synthèse positive à partir de 3 couleurs primaires
- ci-contre, gamut du système **sRGB** dans le diagramme xy



YDbDr, YUV, YIQ, YCbCr...



Autres systèmes colorimétriques basés sur la séparation luminance/chrominance (mais paramètres calculés différemment de Yxy) :

- YDbDr (format analogique SECAM)
- YUV (format analogique PAL)
- YIQ (format analogique NTSC)
- YCbCr (Vidéo numérique, JPEG...)



Image matricielle

Echantillonnage spacial 2D

Image composée de pixels (*picture elements*) généralement carrés

Caractéristiques d'une image :

- définition spatiale
- résolution spatiale
- quantification des couleurs...

Pouvoir de résolution de l'oeil humain

Pouvoir de résolution de l'oeil \approx une minute d'arc (0,017°)

Distance à l'observateur	30cm	3m	300m
Taille d'un détail observable	0,1mm	1mm	10cm

Résolution spatiale d'un périphérique de restitution = densité de pixel (nombre de pixel par unité de surface) :

- **écran**
 - distance entre deux luminophores : de 0.5 à 0.05 mm
 - 3 sous-pixels R-G-B
 - rendu non linéaire (courbe de compensation gamma)
- **imprimante**
 - taille du point coloré : de 0.1 à 0.05 mm (max 720dpi)
 - matrice de sous-pixels YMCK (exemple 8x8) pour tramage

Définition spatiale d'une image

Définition spatiale = nombre de pixels d'une image

type	exemple	largeur (pixels)	hauteur (pixels)	ratio	définition	Mpixel
1er capteur	Kodak '75	100	100	1/1	10.000	0,01
capt. smartphone	iPhone	4.000	3.000	4/3	12.000.000	12
capt. compact	Canon G9x2	5.472	3.648	4/3	19.961.856	20
capt. réflex	Canon 5D4	6.720	4.480	3/2	30.105.600	30
capt. MF	Has. H6D	11.600	8.700	4/3	100.920.000	100
scanner photo	Has. X1	31.500	25.200	5/4	793.800.000	~800
scanner à plat	1200dpi	12.000	9.600	5/4	115.200.000	~100
scanner à plat	2400dpi	24.000	19.200	5/4	460.800.000	~400
scanner à plat	Epson V850	96.000	38.400	5/4	3.686.400.000	3.600

Résolution spatiale d'une image

type	dim. horizon.		ppcm	dpi	distance cm	résolution angulaire
	pixels	cm				
smartphone	480	7	69	174	20	0,042°
Retina iPh.7	1136	10.4	128	326	20	0,022°
QuadHD LG G3	2560	12	212	538	20	0,014°
portable 14'	1366	30	46	116	40	0,031°
Retina 15'	2880	33	87	222	40	0,016°
moniteur 24'	1920	52	37	94	60	0,026°
CinemaHD	2560	58	44	112	60	0,022°
iMac 5K	5 120	58	88	224	60	0,011°
TVSD 30'	720	50	14	37	300	0,013°
TVHD 40'	1920	88	22	55	300	0,009°
TVUHD 50'	3840	122	31	80	300	0,006°
photo 10x15		15,0	118	300	20	0,024°
texte A4		29,7	236	600	40	0,006°
photo A2		59,4	100	256	100	0,006°

Nombre de bits par pixels



nb bits	nb couleurs	remarque
4	16	couleurs indexées (ex : EGA)
8	256	indexées (palette réduite du web)
15/16	35K	<i>High color</i> , 5 bits/composantes
24/32	16M	<i>True color</i> , 8 bits/composantes
48/64	281T	<i>Deep color</i> , 16 bits/composantes

Codages des images

Codage réversible

- codage + décodage : fonction **identité**
- méthodes de compression classique (Huffman, Fano-Shannon, LZW...)
- exemples :
 - **GIF** : palette jusqu'à 256 couleurs, compression LZW
 - **PNG** : palette ou RGB (8 bits/comp.), compression Huffman + LZ77
 - **TIFF** : palette, RGB, CMYK ou Lab (8, 16 ou 32 bits/comp.), LZW
- taux de compression dépendant fortement de l'image

Codage non réversible

- codage : fonction **non injective** (même codage pour 2 images \neq)
- but : même codage pour 2 images **perceptivement** semblables
- nombreuses approches mais la plus utilisées est JPEG

Réduction en profondeur

Possibilité de réduire le nombre de bit par pixels :

- **diminution du nombre de bit par composante**
ex : 16 bits réduit à 8 bits par composantes RGB
▮ baisse de la qualité
- **définition de palettes** (indexage des couleurs)
ex : extraction des 65K couleurs utilisées d'une image puis codage de la référence de ces couleurs (nécessite la transmission de la palette associée) ▮ augmentation de la complexité

JPEG version de base

Joint Photographic Experts Group

Norme ISO/CEI 10918-1 UIT-T Recommendation T.81 (1992)

définit le format d'enregistrement et l'algorithme de décodage pour une représentation numérique compressée d'une image fixe

But : dépasser les compressions sans pertes insuffisantes pour le stockage ou la transmission d'images

- basé sur deux observations :
 - **contenu relativement stable** (peu d'altérations dans une petite zone)
 - l'œil humain est beaucoup **plus sensible aux modifications des basses fréquences** que des hautes fréquences (petits détails et bruit)
- compression basée sur la suppression des hautes fréquences
- choix du taux de compression
- performance de l'ordre de 10 à 100

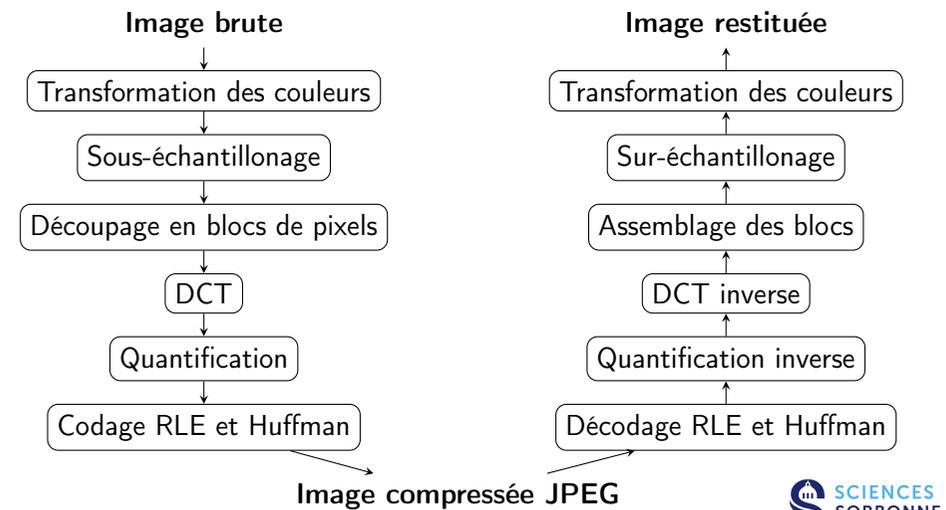
JPEG : exemple



Exemple avec une image originale (RGB 24 bits non compressée) : 230Ko

- image de qualité supérieure : 49Ko
- image de qualité moyenne : 14Ko
- image de qualité inférieure : 7Ko

Processus de codage JPEG



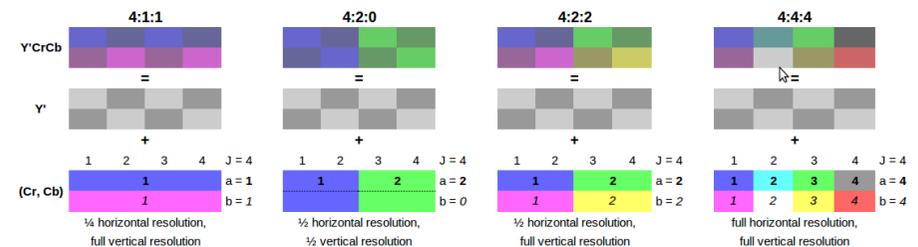
Transformation des couleurs

Passage de l'espace colorimétrique RGB 8 bits à YCbCr 8 bits pour permettre un sous-échantillonnage de la chrominance :

- $Y = 0.299R + 0.587G + 0.114B$
- $Cb = -0.169R - 0.331G + 0.5B + 128$
- $Cr = 0.5R - 0.419G - 0.081B + 128$

- $R = Y + 1.402(Cr - 128)$
- $G = Y - 0.344(Cb - 128) - 0.714(Cr - 128)$
- $B = Y + 1.772(Cb - 128)$

Sous-échantillonnage de la chrominance

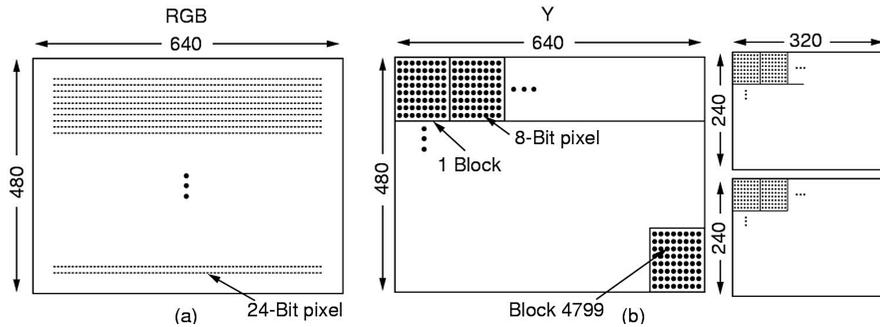


Plus grande sensibilité humaine à la luminance qu'aux différences de couleurs (valeurs usuelles) :

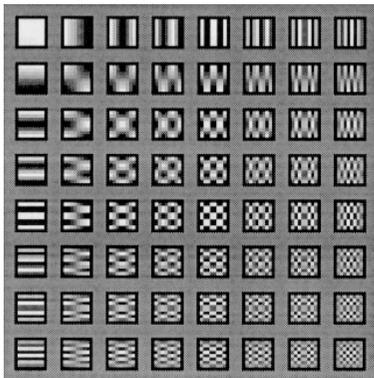
- 4:4:4 YCbCr/RGB (absence de sous-échantillonnage)
- 4:2:2 (réduction horizontale de 1/2 de la taille des blocs Cb/Cr)
- 4:2:0 (réduction horizontale et verticale)

Découpage en blocs 8x8

Exemple pour une image en 640x400 pixels (RGB \rightarrow 4:2:0 YCbCr)



DCT (2)



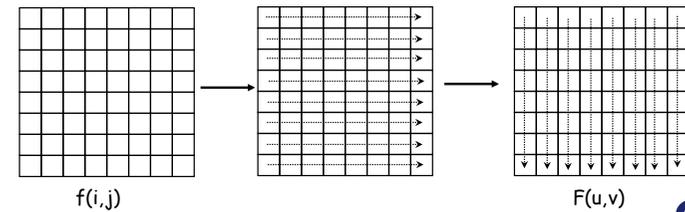
Signification graphique :

- 64 (8 x 8) fonctions de base
- superposition de ces fonctions

DCT (1)

Transformation cosinus discrète :

- conversion d'un bloc 8x8 dans le domaine fréquentiel :
 - la DCT nécessite une matrice carrée
 - une DCT sur l'image entière serait trop coûteux
- calcul DCT bidimensionnelle
 - $F(0,0)$: moyenne du bloc (composante DC)
 - $F(i,j)$: puissances spectrales pour chaque fréquence (coefficients AC)

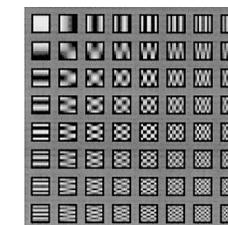


DCT (3)



0 - noir
255 - blanc

700	90	100	0	0	0	0	0
90	0	0	0	0	0	0	0
-89	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0



Quantification

But :

- compression en diminuant la précision des valeurs de la DCT
- précision plus faible pour les haute fréquences

Matrice de quantification :

- division et arrondis des valeurs de la DCT par celles de la matrice de quantification
- coefficients de la matrice de plus en plus grands lorsqu'on s'éloigne de la position (0,0) pour filtrer les hautes fréquences
- JPEG : deux matrices pour chrominance et luminance.

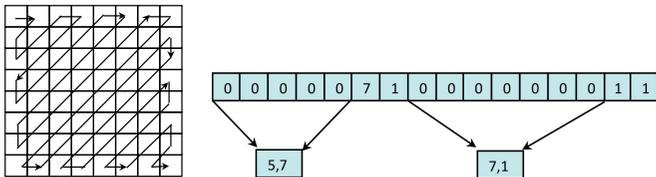
Décompression :

- multiplication par le contenu de la matrice de quantification

Exemple de quantification

Matrice de pixels d'entrée								Matrice DCT quantifiée							
140	144	147	140	140	155	170	175	403	-4	2	-1	2	-1	-1	-1
144	152	140	147	140	148	167	170	4	-5	3	-1	-1	1	1	0
152	155	136	167	163	162	152	172	-1	-3	0	0	-1	0	-1	0
168	145	156	160	152	155	136	160	-1	0	1	-1	0	0	0	0
162	148	156	148	140	136	147	162	0	1	1	0	-1	1	1	1
147	167	140	155	155	140	136	162	0	0	-1	0	0	0	0	0
136	156	123	167	162	144	140	147	1	0	0	0	0	0	0	0
148	155	136	155	152	147	147	136	0	0	0	0	0	0	0	0
Matrice DCT								Matrice DCT déquantifiée (décompression)							
1210	-18	15	-9	23	-9	-14	-10	1200	-20	14	-9	22	-13	-15	-17
21	-34	26	-9	-11	11	14	7	20	-35	27	-11	-13	15	17	0
-10	-24	-2	6	-18	3	-20	-1	-7	-27	0	0	-15	0	-10	0
-8	-5	14	-15	-8	-3	-3	8	-9	0	13	-15	0	0	0	0
-3	10	8	1	-11	18	18	15	0	13	15	0	-10	21	23	25
4	-2	-18	8	8	-4	1	-7	0	0	-17	0	0	0	0	0
0	1	-3	4	-1	-7	-1	-2	15	0	0	0	0	0	0	0
0	-8	-2	2	1	4	-6	0	0	0	0	0	0	0	0	0
Matrice de quantification								Matrice de pixels de sortie (décompression)							
3	5	7	9	11	13	15	17	142	143	154	141	133	153	170	170
5	7	9	11	13	15	17	19	130	152	120	151	144	154	163	181
7	9	11	13	15	17	19	21	150	156	130	166	162	163	154	172
9	11	13	15	17	19	21	23	163	145	160	153	151	153	145	154
11	13	15	17	19	21	23	25	168	150	156	145	140	130	141	150
13	15	17	19	21	23	25	27	148	164	133	164	158	140	136	163
15	17	19	21	23	25	27	29	130	150	123	164	165	140	134	145
17	19	21	23	25	27	29	31	148	156	140	148	150	146	153	141

RLE et Huffman



Dernières étapes du codage JPEG :

- parcours en zig-zag du bloc 8x8 (regroupe les hauts coefficients en début de vecteur)
- RLE ((Run-length encoding)) :
 - les hautes fréquences sont en principe presque toutes nulles
 - longues séquences de 0
 - remplace le vecteur par une suite (Skip, Value) :
 - Skip = nombre de 0 consécutifs
 - Value = valeur suivant le dernier 0 (nb bits pour coder + valeur)
- codage Huffman standard

JPEG : évolutions

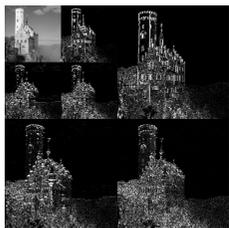
Besoins identifiés :

- amélioration de la performance de compression
- réduction des artefacts pour les taux de compressions élevés (bloc 8x8 visibles)
- multi-résolution
- transmission progressive
- compression variable selon la zone de l'image (ROI : région d'intérêt)
- support de tous espaces de couleurs
- toutes profondeurs de bit par pixels (entière et réelle)
- prise en compte de canaux supplémentaires (alpha/transparence)
- résistance aux erreurs

JPEG2000

Norme ISO/CEI 15444-1 UIT-T Rec. T.800 (2000)

De manière identique à JPEG, définit uniquement le format d'enregistrement et l'algorithme de décodage



(CC BY-SA 3.0 Alessio Damato)

But : améliorer la compression JPEG en utilisant une approche basée sur les ondelettes (au lieu de la DCT)

- processus de codage similaire à JPEG :
 - transformation en couleur : YUV ou YCbCr
 - sous échantillonnage
 - **découpage en sous bande**
 - **transformée en ondelettes**
 - quantification
 - codage entropique spécifique

CONT : plan du cours 2/5

- 1 Introduction
 - contenus numériques
 - encodage des textes
 - encodage des images
- 2 Contenus audio
 - audio numérique
 - encodages sans pertes
 - compressions avec pertes
- 3 Contenu vidéo
 - vidéo numérique
 - encodages sans pertes
 - compressions avec pertes

CONT : plan du cours 2/5

- 1 Introduction
 - contenus numériques
 - encodage des textes
 - encodage des images
- 2 Contenus audio
 - audio numérique
 - encodages sans pertes
 - compressions avec pertes
- 3 Contenu vidéo
 - vidéo numérique
 - encodages sans pertes
 - compressions avec pertes

Transmission de contenu audio

Dépend de l'usage du contenu audio :

- **téléchargement antérieure à la lecture** du fichier de contenu
 - systèmes de partage de fichiers (ex : podcast)
- **lecture pendant le téléchargement** (*streaming*)
 - diffusion à la demande de fichiers pré-enregistrés (audio pour le web)
 - fonctionnalités de lecteur (avancer, reculer, etc.)
- **lecture de contenu produit en direct** (*live streaming*)
 - diffusion d'émissions en *live* (web radio)
- **échanges de contenu interactifs**
 - VoIP
 - musique collaborative

4 niveaux d'interactivité

Impact de l'interactivité :

Non interactif \Rightarrow plus de temps pour coder/décoderInteractif \Rightarrow codage/décodage rapide

- non-interactif avec pré-chargement
 - codage long et décodage long possibles
- non-interactif avec téléchargement à la demande
 - codage long possible
 - buffer réduit en réception : décodage rapide
- non-interactif avec téléchargement de contenus en direct
 - latence réduite : codage rapide
 - buffer réduit en réception : décodage rapide
- interactif
 - codage et décodage très rapides
 - contraintes de type télécom

Formats audio courants

Musique :

- PCM, CD audio...
- FLAC, M4A, MPEG4-ALS, ATRAC/WMA Lossless...
- MP3, AAC, Ogg, ATRAC/WMA Lossy...

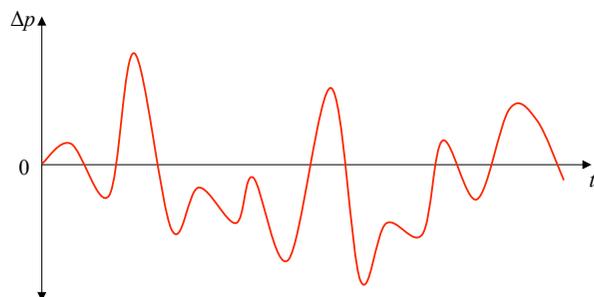
Voix :

- A-law et μ -law (G711)
- GSM, AMR-WB (G.722.2)...

 \Rightarrow importance de la compression

- formats bruts (PCM...)
- compression sans perte (Huffman, LZW...)
- compression avec perte (approches psycho-acoustiques)
 - basées sur les caractéristiques de l'audio et de l'ouïe
 - très efficace mais destructif

Signal audio analogique

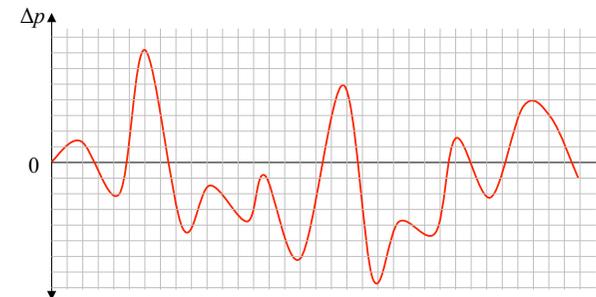
 t : temps

- unité la plus fine (typique) : μs
- échelle : jusqu'à plusieurs heures

 Δp : pression relative

- unité la plus fine (typique) : $20 \mu\text{Pa}$
- échelle : jusqu'à environ 100 Pa

Numérisation du signal audio

Grille d'analyse **discrete** du signal audio selon :

- l'axe des temps (linéaire)
- l'amplitude du signal (pas forcément linéaire)

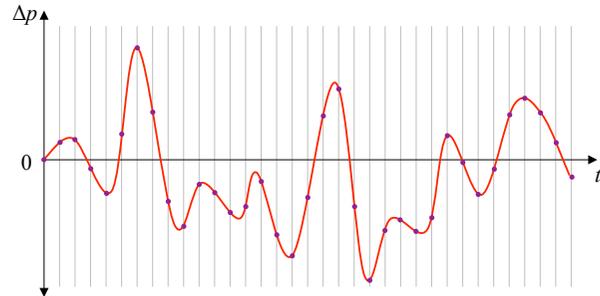
 \Rightarrow introduit 2 étapes :

- échantillonnage
- quantification

Echantillonnage du signal audio

Théoreme de Nyquist-Shannon

La représentation discrète d'un signal de fréquence maximale F_{max} par des échantillons régulièrement espacés exige une fréquence d'échantillonnage supérieure à $2F_{max}$.



Spectre des fréquences vocales

Gamme usuelle de 500 à 2000 Hz

- variable selon les individus 80 à 3800 Hz



Appelée "bande étroite" (*narrow-band*)

- canal audio ITU : 300 à 3400 Hz
- bande passante de 3,1 kHz

➡ fréquence d'échantillonnage $> 6,8$ kHz

- exemple de la téléphonie :
 - échantillonnages ITU à 8 kHz (synchronisation de tout le réseau RTC)
 - 1 échantillon toutes les $125\mu s$

Spectre des fréquences audibles

Gamme de 20 Hz à 20 kHz

- dépend des individus et de l'âge



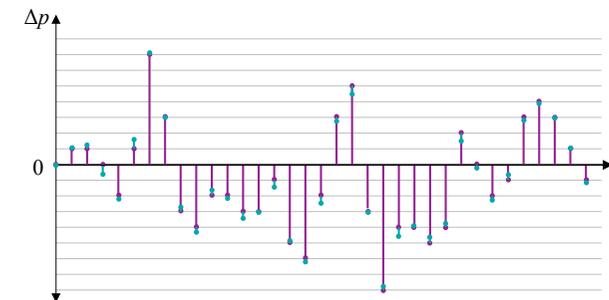
Appelée "large bande" (*wide-band* ou *full-band*)

- presque 20 kHz

➡ fréquence d'échantillonnage > 40 kHz

- 1 échantillon toutes les $25\mu s$
- exemple de support numériques :
 - CD-audio : 44,1 kHz
 - DVD-audio : jusqu'à 192 kHz

Quantification du signal audio



Procédé qui permet d'approcher un signal continu par les valeurs d'un ensemble discret

- opération destructrice d'information
- introduction d'une erreur (ou bruit) entre le signal quantifié et le signal source

Etendue dynamique

Equivalence rapport/dB

rapport r	$10 \log_{10}(r)$ dB
1,00	0 dB
1,26	1 dB
1,58	2 dB
2,00	3 dB
2,51	4 dB
3,16	5 dB
3,98	6 dB
5,01	7 dB
10,00	10 dB
20,00	13 dB
100,00	20 dB
1000,00	30 dB

Rapports entre deux valeurs :

- la puissance maximum du signal
- la puissance minimale décelable du signal
- pour une quantification sur m bits : $\frac{1}{2^m}$

Le niveau sonore est exprimé en décibel (dB)

- indique le rapport puissance
- lié à une valeur de référence

Etendue dynamique humaine

Equivalence rapport-dB

0 dB	limite audible
20-30 dB	chuchotement
40-50 dB	bibliothèque
60-70 dB	télévision
70-80 dB	aspirateur
90-100 dB	circul. dense
110-120 dB	discothèque
120 dB	seuil douleur
130 dB	décollage avion
180 dB	décollage fusée

Sensibilité à de l'intensité sonore :

- dédoublement (triplement, etc.)
- division par deux (trois, etc.)

Définition de la référence **0dB SPL** (*Sound Pressure Level*) :

- intensité acoustique : 1 pW/m^2
- pression acoustique : $20 \text{ } \mu\text{Pa}$

Seuil de sensation : 200 Pa

- au dessus, non-supportable

Dynamique de l'oreille humaine :

- rapport de pression de 10^7
- rapport de puissance de 10^{14}
- échelle log. : **140 dB**

Etendue dynamique des CD/DVD

CD : qualité suffisante ?

- **16 bits**/échantillon
- $20 \log_{10}(2^{16}/1) = \mathbf{96 \text{ dB}}$



Comparaison des étendues dynamiques :

- orchestre : 90 dB
- humaine : 140 dB

DVD-Audio : augmentation de l'étendue dynamique

- jusqu'à **24 bits**/échantillon
- $20 \log_{10}(2^{24}/1) = \mathbf{144 \text{ dB}}$

Nombreux encodages audio

Codages numériques bruts

- large bande (tout le spectre audible) :
 - MIC/PCM, CD audio, DVD audio, DSD...
- bande étroite (voix) :
 - A-law et μ -law (G711)

Codages avec compression non destructive

- large bande :
 - FLAC, MA4, MPEG4-ALS, ATRAC/WMA Lossless

Codages avec compression destructives

- large bande :
 - MP3, AAC, Ogg, ATRAC/WMA Lossy
- bande étroite :
 - GSM, AMR-WB (G.722.2)

Codec/conteneur

Attention, certains formats de fichiers audio n'incluent pas directement un codage mais sont des structures permettant d'intégrer différents éléments :

- métadonnées
- multiples codages
- différents médias (plusieurs pistes audio, vidéo...)
- mécanismes de multiplexage (découpage en atomes, blocs, chunks, packets ou segments)

► un lecteur audio doit être capable d'ouvrir un fichier "conteneur" et décoder les différents codecs nécessaires pour les flux qu'il intègre
Exemple de format "conteneur" :

- Exclusivement audio : AIFF (Apple), WAV (Windows), XMF...
- Multimédia : 3GP, AVI (Windows), MOV (QuickTime), Flash, Mkv, MPEG-TS, MP4, OGG, RM, DivX...



CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

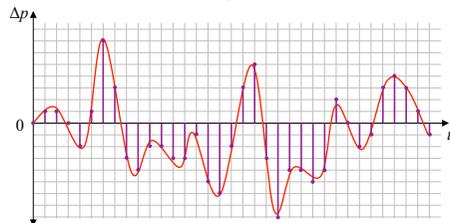
3 Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes



MIC/PCM

Modulation par Impulsion et Codage / *Pulse Code Modulation*



- représentation numérique non compressée d'un signal analogique
- basé sur un échantillonnage et quantification
 - codage linéaire : CD audio, DVD audio...
 - codage logarithmique : A-law, μ -law...
- optimisations : modulation différentielle (DPCM), adaptative (ADPCM)



AES3 (AES/EBU)

Standard de communication pour l'audio numérique (1985) :



- 2 canaux audio 20/24 bits à 44.1 et 48 kHz
 - évolution : 88.2, 96, 176.4 ou 192 kHz
- interface professionnelle
 - câble torsadé à 3 conducteurs symétriques 110 Ω : interface XLR (IEC 60958 Type I)
 - câble coaxial 75 Ω : connecteur BNC (AES 3id)
- interface grand public : S/PDIF
 - câble coaxial 75 Ω : connecteur RCA (IEC 60958 Type II)
 - fibre optique : connecteur Toslink (IEC 60958 Type II Optical)
- profils RTP : L20 et L24 (RFC 3190)



CD audio

COMPACT
disc
DIGITAL AUDIO**Support** commercialisé en 1982

- 2 canaux audio (stéréo)
- échantillonnage à 44100 Hz
- quantification linéaire sur 16 bits

Débit correspondant

- 1 canal (mono) : ~ 700 kbit/s,
- 2 canaux (stéréo) : ~ 1400 kbit/s,

Profil RTP **L16** (RFC 3551)

- horloge RTP : 44100
- taille de paquet par défaut : 20ms



DVD audio

DVD
AUDIO™**Support** commercialisé en 2000

- 1 à 6 canaux : mono (1.0), stéréo (2.0 et 2.1), stéréo + mono-surround ou 3-stéréo (3.0 et 3.1)... full-surround (5.0 et 5.1)
- échantillonnage à 44.1, 48, 88.2, 96, 176.4 ou 192 kHz
- quantification linéaire sur 16, 20 ou 24 bits

Débit correspondant

- stéréo 192/24 : ~ 8.8 Mbit/s,
- 5.1 96/24 : ~ 13.2 Mbit/s ➡ compression sans perte (débit max 9.6 Mbit/s)

Profils RTP **L16**, **L20** et **L24** (RFC 3190)

Autres encodages audio sans pertes (1)

SACD commercialisé en 1999 (concurrent du DVD audio)

- jusqu'à 6 canaux avec compression sans perte
- échantillonnage DSD (**Direct Stream Digital**) :
 - 1 bit à 2.8224 MHz (64x la fréquence du CD)
 - dynamique supérieure à 120dB
 - bande passante étendue jusqu'à 100kHz
 - ~ quantification sur 20 bits



SUPER AUDIO CD

**Blu-Ray** commercialisé en 2006

- 8 canaux à 48/96 ou 6 à 192 kHz, plusieurs formats :
 - **PCM** linéaire
 - 16, 20 ou 24 bits, débit max. : ~ 28 Mbit/s
 - Dolby **TrueHD** (suite AC-3)
 - 16 ou 24 bits, débit max. : ~ 18 Mbit/s
 - **DTS HD Master Audio** (combo avec et sans pertes)
 - 16 ou 24 bits, débit max. : ~ 24 Mbit/s



Autres encodages audio sans pertes (2)

- **FLAC** (*Free Lossless Audio Codec*)
 - créé en 2001, incorporé par Xiph.org en 2003
 - compression LPC de 1 à 8 canaux (4 à 32 bits jusqu'à 655 KHz)
 - CRC pour identifier les trames corrompues
- **ALAC** (*Apple Lossless Audio Codec*)
 - créé en 2004, libre en 2011,
 - également compression LPC (prédiction linéaire)
- **MPEG4-ALC** (*Audio Lossless Codec*)
 - créé en 2006, intégré à MPEG-4 part 3
 - compression LPC multi-canaux (jusqu'à 32 bits et 384 KHz)
 - version complémentaire à AAC : MPEG4-SLS (*Scalable to Losses*)
- **WMA Lossless**
 - créée en 2003 par Microsoft
 - compression multi-canaux (jusqu'à 24 bits 96 KHz)



ITU-T G-711

Normalisation du **codage MIC** pour la téléphonie numérique (1972)

- échantillonnage à **8 kHz** pour bande passante 300–3400 Hz
- débit de 64 kbps
- pas de quantification non régulier (échelle logarithmique)
- deux algorithmes de quantification :
 - loi μ (μ -law) aux USA
 - loi A (A-law) ailleurs
- profils RTP : **PCMU** et **PCMA** (RFC 3551)
- extensions :
 - **G.711.0** ou G.711 LLC (*Lossless Data Compression*)
 - **G.711.1** (*Wideband Embedded Extension*)
 - échantillonnage bande étroite (8 kHz) ou large (16 kHz)
 - bande passante de 50 à 4000 Hz ou 7000 Hz
 - débits de 64, 80 ou 96 kbit/s

A-law, μ -law

Deux encodage à intervalles de quantification irrégulier sur 8 bits :

Codage A-Law :

entrée 13 bit MIC linéaire

Linear code	A-law code
s0000000wxyz'a	s000wxyz
s0000001wxyz'a	s001wxyz
s000001wxyz'ab	s010wxyz
s00001wxyz'abc	s011wxyz
s0001wxyz'abcd	s100wxyz
s001wxyz'abcde	s101wxyz
s01wxyz'abcdef	s110wxyz
s1wxyz'abcdefg	s111wxyz

~ nombre flottant avec 1 bit de signe,
4 bits de mantisse et 3 bits d'exposant

Codage μ -Law

entrée 14 bit MIC linéaire + 32

Linear code	Mu-law code
s00000001wxyz'a	s000wxyz
s0000001wxyz'ab	s001wxyz
s000001wxyz'abc	s010wxyz
s00001wxyz'abcd	s011wxyz
s0001wxyz'abcde	s100wxyz
s001wxyz'abcdef	s101wxyz
s01wxyz'abcdefg	s110wxyz
s1wxyz'abcdefgh	s111wxyz



CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

3 Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes



Compression avec pertes

Réduire fortement le débit sans trop dégrader la qualité sonore :

- traitements avancés des signaux (utilisation de DSP)
- compromis entre plusieurs caractéristiques
 - débit
 - qualité (introduction du MOS, *Mean Opinion Score*, de 1 à 5)
 - retard (introduit par la compression)
 - complexité de l'algorithme de compression (MIPS nécessaire)



Psycho-acoustique

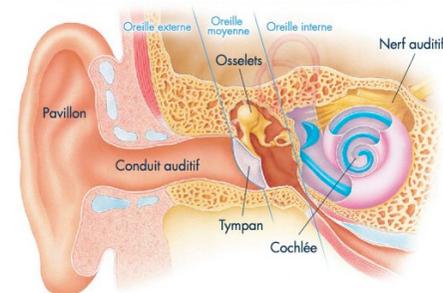
Codages audio : prise en compte des capacités de l'ouïe :

- transformation en représentation fréquentielle
- suppression de ce qui n'est pas perçu
 - sensibilité variable en fonction de la fréquence
 - masquage des sons entre eux

Codages vocaux : prise en compte des capacités la voix et de sa perception :

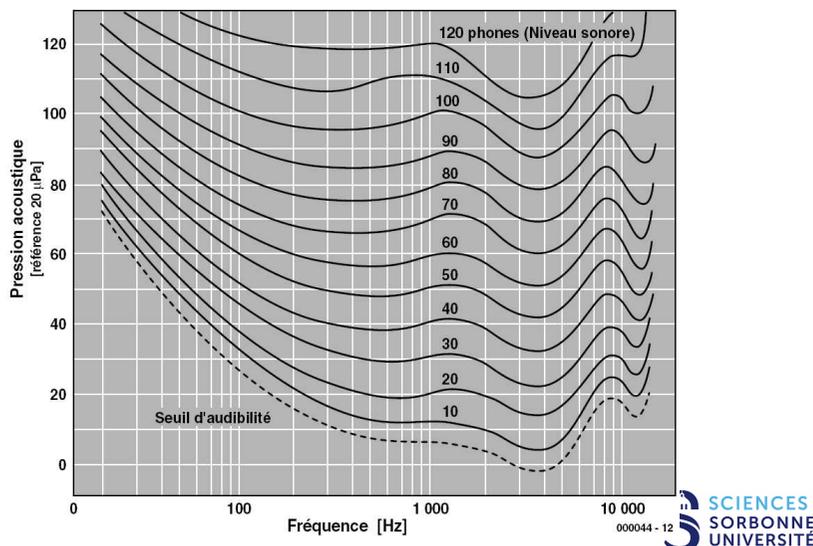
- restitution de la voix par synthèse
 - modélisation du canal vocal

Tonalité audible

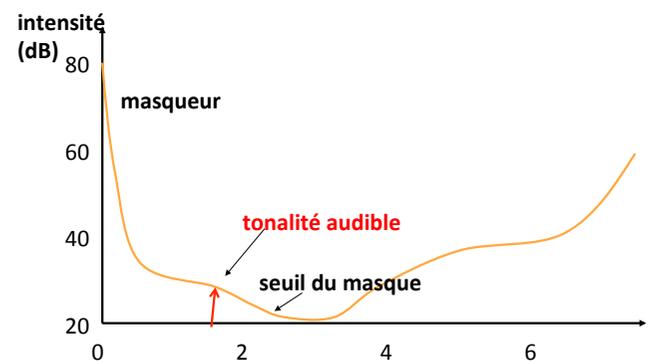


- transformation du son en vibrations grâce au tympan et aux osselets
- vibrations transmises à un fluide situé dans la cochlée
- transmission au nerf auditif par des cellules ciliées
 - une membrane de consistance variable est étirée dans la cochlée (indique la hauteur tonale sur un spectre de résonance déterminée)

Sensibilité auditive humaine



Tonalité audible



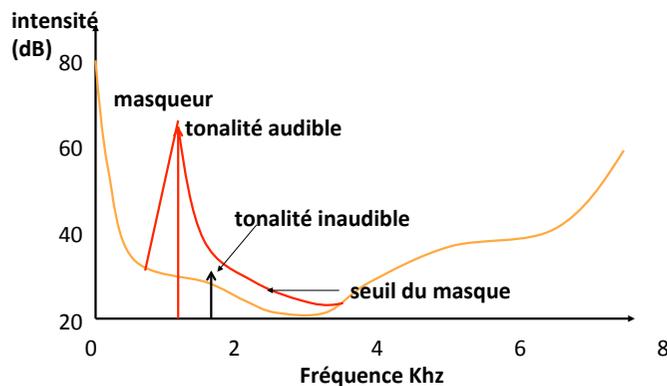
- spectre audible compris entre 20 Hz et 15 kHz (en moyenne)
 - **seuil d'audibilité** = niveau minimum d'une tonalité pure détectable dans un environnement **silencieux**

Bandes critiques

Band No.	Center Freq. (Hz)	Bandwidth (Hz)	Band No.	Center Freq. (Hz)	Bandwidth (Hz)	Band No.	Center Freq. (Hz)	Bandwidth (Hz)
1	50	-100	10	1175	1080-1270	19	4800	4400-5300
2	150	100-200	11	1370	1270-1480	20	5800	5300-6400
3	250	200-300	12	1600	1480-1720	21	7000	6400-7700
4	350	300-400	13	1850	1720-2000	22	8500	7700-9500
5	450	400-510	14	2150	2000-2320	23	10,500	9500-12000
6	570	510-630	15	2500	2320-2700	24	13,500	12000-15500
7	700	630-770	16	2900	2700-3150	25	19,500	15500-
8	840	770-920	17	3400	3150-3700			
9	1000	920-1080	18	4000	3700-4400			

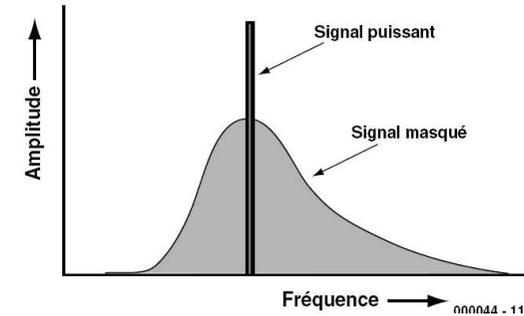
- comportement sélectif de la membrane basilaire
- l'ouïe subdivise le spectre des fréquences audibles en 25 sous-bandes
 - les sous-bandes sont appelées **bandes critiques** (ou *barks*)
 - chaque bande est définie par sa fréquence centrale et sa largeur

Masquage (2)



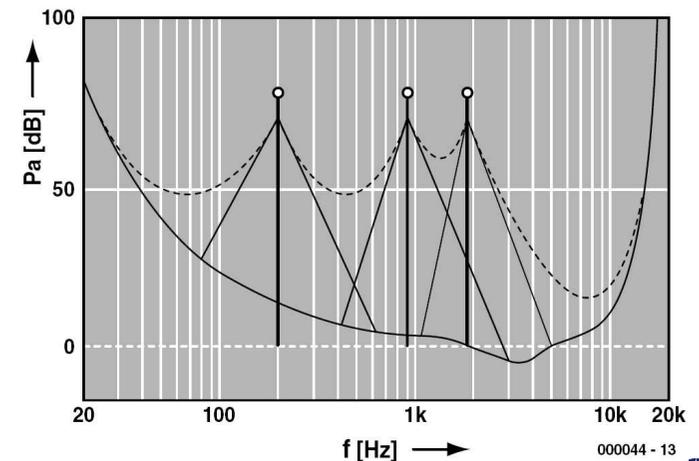
- sur le spectre audible compris entre 20 Hz et 15 kHz
 - **seuil de masquage** = seuil d'audibilité dans le silence + zones inaudibles à cause des masqueurs

Masquage (1)



- **masqueur** :
 - une tonalité puissante centrée sur une bande critique sature la zone concernée
 - une autre tonalité moins puissante située dans la même bande ne sera pas perçue → masquage par la tonalité principale

Masquage (3)



MPEG Audio Layer 3

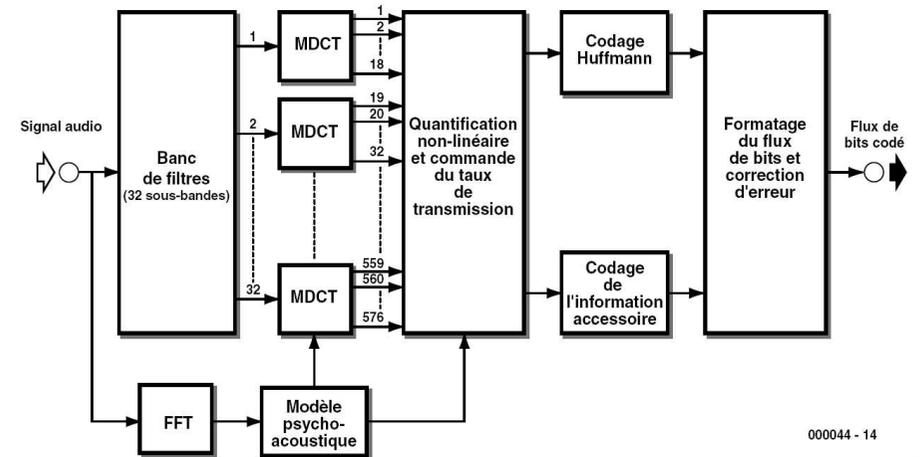
Définis par le sous-groupe Audio de MPEG :

- MPEG-1 Audio (part 3) avec les *Layer I, II et III* (ISO/IEC 11172-3 :1993)
- MPEG-2 Audio (part 3) avec extensions (ISO/IEC 13818-3 :1995)

Caractéristique de l'encodage MP3 (+ extensions MPEG-2) :

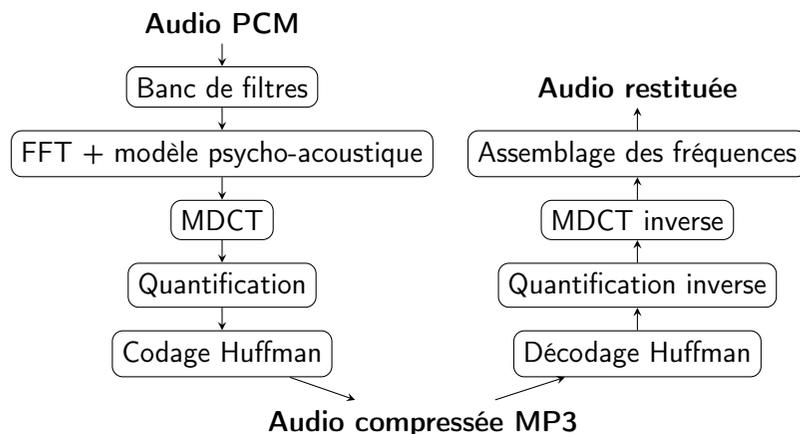
- entrée : PCM 16 ou 24 bits à 48, 44.1, 32 (+ 22.1 et 16 kHz)
 - prise en compte des modes mono, stéréo et double mono (+ 5.1)
- sortie : par voies **160 à 32 Kbps** (+ 16 et 8 Kbps)
 - approche de type **codage perceptuel**
 - efficacité optimisé pour 1.33 bits/échantillons
 - accès aléatoire
 - possibilité de détection d'erreur
 - adaptation au **streaming**

Shéma fonctionnel d'un encodeur MP3



000044 - 14

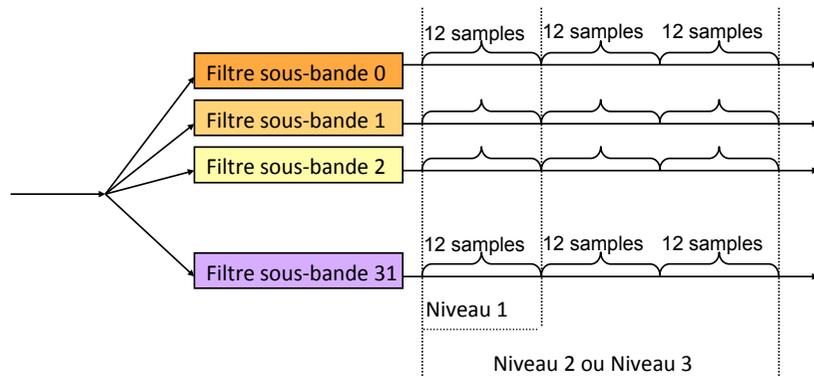
Processus de codage MP3



Mécanismes de bases MP3

- regroupement en trame de 1152 échantillons (18, 26 ou 36 ms)
- banc de 32 filtres
 - différents des bandes-critiques mais simples
 - ré-amplification par bandes
- FFT et utilisation du modèle psycho-acoustique
 - FFT réalisée sur 1024 point
 - identification des masqueurs
 - déduction de la courbe de masquage
- MDCT (*Modified Discrete Cosinus Transform*)
 - découpe en 18 sous-sous-bandes → 576 bandes-étroites (28 à 55 Hz)
 - permet de s'approcher des bandes-critiques
- quantification non linéaire
 - suppression des composantes totalement masquées
- codage de Huffman

Traitement par sous-bande



Exemple de spectre divisé en 32 sous-bandes (divisées elles-même en fenêtres)

Formatage du flux MP3

Flux organisé en trames indépendantes :

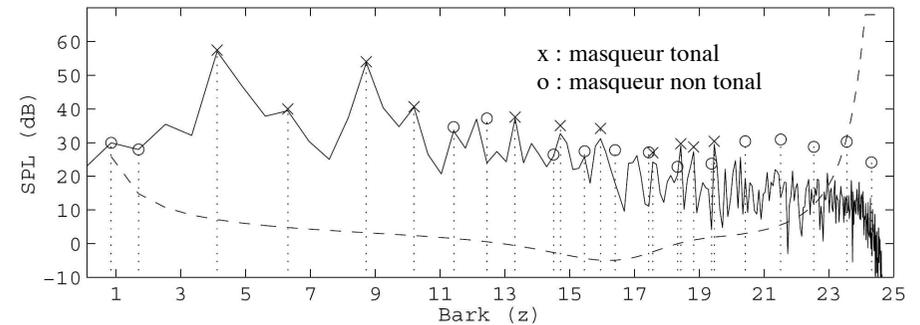
- correspondent 1152 échantillons (18, 26 ou 36 ms)
- contiennent toutes les info nécessaire au décodage
 - entête de 32 bits (débit, échantillonnage...)
 - protection (optionnel)
 - coefficients d'amplification
 - paramètres de quantification
 - données audio compressées

Profil RTP : **MPA** avec une horloge à 90000 (RFC 3551)

Internet Media type :

- audio/mpeg
- audio/mpa et audio/mpa-robust

Application du modèle psycho-acoustique



Shémas de Marcel Cremmel

- avec la FFT, pour chaque bande-critique, identification de :
 - **masqueur tonal** (le ton de niveau maximum si le niveau des tonalités voisines < 7 dB)
 - **masqueur non-tonal** (combinaison de tous les tons non-masqueurs et création d'un masqueur virtuel centré sur la bande-critique)

AAC

Définis par le sous-groupe Audio de MPEG :

- MPEG-2 Audio (part 7) (ISO/IEC 13818-7 : 1997)
- MPEG-4 Audio (part 3.4) (ISO/IEC 14496-3 : 1999)

Successeur du MP3, utilisé pour :

- HDTV en Europe (DVB) et au Japon (ISDB-T)
- iTunes
- téléphones mobiles actuels

Amélioration par rapport au MP3 :

- entrée : plus de fréquences : de 8 à 96 kHz
 - prise en compte jusqu'à 48 canaux
- sortie : débit au choix, taille de trame variable
 - trame de longueur variable (adaptation aux transitoires)
 - MDCT unique (pas de filtrage préalable) ou multiple (8)
 - nombreux module additionnels, profils...

- évolution de la qualité perçue difficile à mesurer

▀ **dépend de l'encodage**

Encodage modulaire AAC

MPEG-2 Part 7 (Advanced Audio Coding) :

- Low Complexity (LC), le plus simple et répandu
- Main Profile (Main), idem LC avec prédiction antérieure
- Scalable Sample Rate(MPEG-4 AAC-SSR), échantillonnage variable

MPEG-4 Part 3 standard (MPEG-4 Audio) :

- Main Audio Profile (1999), utilise les objets de base MPEG-4
- Scalable Audio Profile (1999), AAC-LC, AAC-LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI
- Speech Audio Profile (1999), codec vocaux CELP, HVXC, TTSI
- Synthetic Audio Profile (1999), synthèse TTSI
- High Quality Audio Profile (2000), AAC-LC, AAC-LTP, AAC Scalable, CELP, ER-AAC-LC, ER-AAC-LTP, ER-AAC Scalable, ER-CELP
- Low Delay Audio Profile (2000), CELP, HVXC, TTSI, ER-AAC-LD, ER-CELP, ER-HVXC
- Mobile Audio Networking Profile (2000), ER-AAC-LC, ER-AAC-Scalable, ER-TwinVQ, ER-BSAC, ER-AAC-LD
- AAC Profile (2003), AAC-LC (AAC à faible complexité)
- High Efficiency AAC Profile (2003), AAC-LC, SBR (AAC+)
- High Efficiency AAC v2 Profile (2006), AAC-LC, SBR, PS (eAAC+)
- High Definition AAC Profile (2009), AAC-LC + SLS (complémentaire)



Format du flux AAC

Flux organisé en trames indépendantes :

- taille des échantillons variables
- taille des trames variable :
 - petite taille pour les codes vocaux (63 octets minimum)
 - grande taille pour les codec audio (8191 octets maximum)
 - contiennent toute les info nécessaire au décodage

Internet Media type :

- audio/aac
- audio/aacp
- audio/3gpp
- audio/3gpp2
- audio/mp4
- audio/mp4a-latm
- audio/mpeg4-generic

Profil RTP : **mpeg4-generic**
(RFC 3640)



Codages audio avec pertes des supports optiques

Codec avec pertes DVD-Vidéo commercialisé en 1995



- MP2 (MUSICAM)
 - 1 à 7.1 à jusqu'à 48 kHz, débit max. :~ 900 Kbit/s
- AC-3 (Dolby Digital, ATSC A/52)
 - 1 à 5.1 à 48 kHz, débit max. :~ 450 Kbit/s
- DTS
 - 2.0 à 6.1 à 48 ou 96 kHz, débit max. :~ 1.5 Mbit/s



Codec avec pertes Blu-Ray commercialisé en 2006



- AC-3 (Dolby Digital, ATSC A/52)
 - 1 à 5.1 à 48 kHz, débit max. :~ 640 Kbit/s
- Dolby Digital Plus
 - 1 à 7.1 à jusqu'à 48 kHz, débit max. :~ 4.7 Mbit/s
- DTS Digital Surround
 - 1 à 5.1 à 48 kHz, débit max. :~ 1.5 Mbit/s



GSM (2G)

Norme ETSI EN 300 580 (1990) et 961 (2000)

Codec audio pour la bande vocale

- bande des 300Hz – 3400Hz
- débits de sortie de 13 kbps (*Full Rate*) 6.5 kbps (*Half Rate*)
- technique de codage prédictif linéaire (LPC)
 - block de parole de 20 ms
 - calcul d'autocorrélation
 - filtrage prédictif
 - synthèse modélisant le conduit vocal
- délais de codage de 20 à 30 ms
- evolution vers l'EFR (*Enhanced Full Rate*) à 12.2 kbps en 1997...

Profils RTP : **GSM** et **GSM-EFR** (RFC 3551)



3GPP et 3GPP2

Définition des codecs téléphoniques à partir de la 3G

Codec audio du 3GPP :

- AMR-NB (ou EFR)
- AMR-WB et AMR-WB+
- AAC-LC et HE-AAC (v1 ou v2)

Codec audio du 3GPP2 :

- idem ceux du 3GP
- EVRC, EVRC-B et EVRC-WB
- QCELP
- SMV
- VMR-WB

AMR Voice Codec (téléphonie mobile)

AMR Adaptive Multi-Rate (EFR, GSM-AMR, AMR-NB) (1999)

- codage de la voix (300-3400 Hz)
- trame de 160 échantillons sur 13 bits à 8 kHz (20 ms)
- complexité moyenne : ACELP (LPC avec dictionnaire)
- débits de sortie de 4.75 à 12.2 kbps, délais de 25 ms

AMR-WB Adaptive Multi-Rate Wideband (ITU-T G.722.2) (2002)

- codage de la voix étendue (50-7000 Hz)
- ALCEP sur trame de 20ms (14 bits à 16 kHz)
- débits de sortie de 6.6 à 23.85 kbps, délais de 25 ms

AMR-WB+ Extended Adaptive Multi-Rate Wideband (2004)

- codage étendu jusqu'à 20 kHz
- échantillons sur 16 bits à 16, 24, 32 ou 48 kHz
- débits de sortie de 5,2 à 36 kbps mono, 6.2 à 48 kbps stéréo
- compatibilité ascendante avec AMR, délais de 60 à 90 ms

ITU-T Voice Codec (VoIP)

G.729 Codage à 8kbps (1996)

- codage de la voix (300-3400 Hz)
- CS-ALCEP sur trame de 80 échantillons sur 16 bits à 8 kHz (10 ms)
- débits de sortie de 6.4, 8 ou 11.8 kbps, complexité élevée
- version simplifiée G.729a seulement à 8 kbps, délais de 15 ms

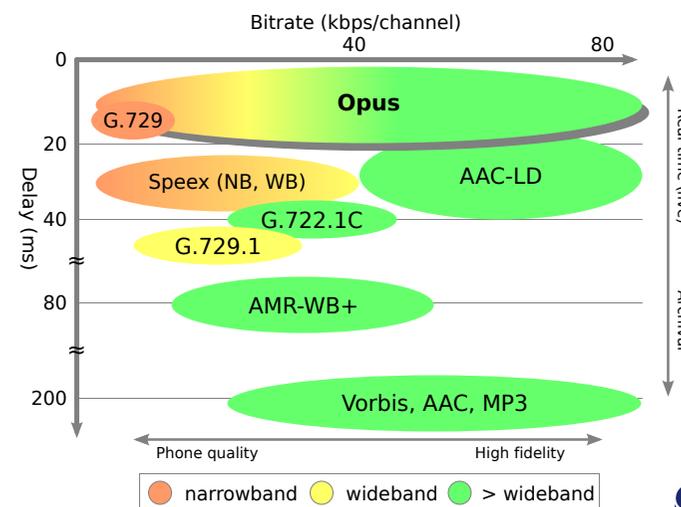
G.722.1 (1999)

- codage de la voix étendue (50-7000 Hz)
- MLT sur trame de 20ms (16 bits à 16 kHz), complexité faible
- débits de sortie de 24 et 32 kbps, délais de 40 ms
- version G.722.1C à 32 kHz, sortie 48 kbps

G.729.1 Extension de G.729 (2006)

- échantillons sur 16 bits à 8 ou 16 kHz
- débits de sortie de 8 à 32 kbps, délais de 40 ms
- Système en couche à partir d'un flux de base G.729

Opus (RFC 6716)



CONT : plan du cours 2/5

① Introduction

contenus numériques
encodage des textes
encodage des images

② Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

③ Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes

Caractéristiques des flux vidéo

- nombre d'image par seconde (24, 25, 30, 50, 60...)
- entrelacé ou progressif (i ou p)
- définition de l'image
- format d'image (4/3, 16/9, 1.85, 2.35...)
- forme des pixels (carré ou rectangulaire)
- espace colorimétrique et sous-échantillonnage
- bits par pixels (8, 10, 12 bits par composantes...)
- codage et compression

CONT : plan du cours 2/5

① Introduction

contenus numériques
encodage des textes
encodage des images

② Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

③ Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes

Télévision analogique

Limités par les capacités de transmission et le tube cathodique (entrelacé)



• noir et blanc

- démarrage dans les années '30,
- transmission du signal **Luma**
- nombreux formats (de 30 jusqu'à 819 lignes)
(~ 737i50 monochrome)

• couleur

- ajout du signal Chroma (années '50),
- 3 principaux formats entrelacés :
 - NTSC : YIQ 525 lignes à 60Hz
(~ 488i60)
 - SECAM : YDbDr 625 lignes à 50Hz
(~ 576i50)
 - PAL : YUV 525 l. 60hz et 625 l. 50hz

Evolution de la télévision analogique

Vers le numériques :

- D-MAC/D2-MAC (multiplexage analogique temporel)
- NICAM (son numérique stéréo) :
 - échantillonnage à 32kHz
 - quantification logarithmique (14 → 9 bits)

Haute définition analogiques :

- **MUSE/HiVision** (Japan)
 - 1125 lignes entrelacée (~ 1035i60)
- **HD-MAC** (Europe)
 - 1250 lignes entrelacées (~ 1152i50)
 - abandonné en 93 avec l'apparition du numérique (DVB-T)

Arrêt progressif des services analogiques (ex : 2010 en France)



Formats vidéo numériques

SDTV : 480i60 (NTSC), 480p30, 576i50 (PAL/SECAM) et 576p25

EDTV : 480p60 et 576p50

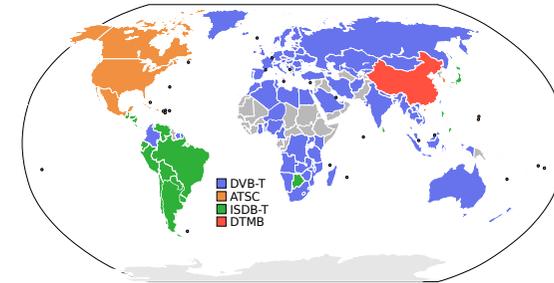
HDTV : 720i50, 720i60, 720p*, 1080i50, 1080i60 et 1080p*

UHDTV : 2160p* et 4320p*

- définitions
 - SD : systèmes conventionnels (NTSC, PAL/SECAM)
 - ED (Enhanced Def) : mode progressif (DVD)
 - HD (HD Ready, Full HD) : haute définition (SD*2)
 - UHD (4K et 8k) : ultra haute définition (HD*2 ou HD*4)
- mode entrelacé ou progressif :
 - systèmes conventionnels (NTSC, PAL, SECAM) sont entrelacés
 - DVD, HDTV, UHDTV et moniteurs informatiques sont progressifs
- rafraîchissement usuels : 24, 25, 30, 50 ou 60 images par seconde
- ratio : 16/9 ou 4/3



Télévision numérique



(CC BY-SA 3.0 EnEdC)

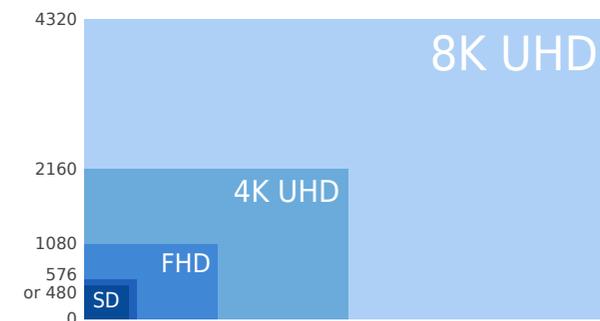
4 formats de transmission terrestre :

- **DVB-T** (initiative européenne à standard ouvert)
- **ATSC** (US, remplaçant numérique du NTSC)
- **ISDB-T** (Japon, version numérique de MUSE)
- **DTMB** (Chine)

Tous en H.262/MPEG-2 Part 2 et H.264/MPEG-4 AVC (sauf DTMB)



Poids d'un signal vidéo (1)



Débit SD : $768 \times 576 \times 25$ images/s ~ 11 millions de pixels/s

Débit HD : $1920 \times 1080 \times 25$ images/s ~ 52 millions de pixels/s

Débit 4K : $3840 \times 2160 \times 25$ images/s ~ 210 millions de pixels/s

Débit 8K : $7680 \times 4320 \times 25$ images/s ~ 830 millions de pixels/s



Poids d'un signal vidéo (2)

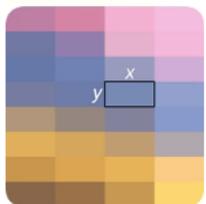
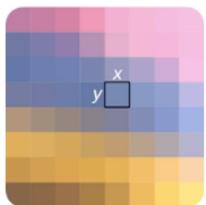
Quantification RGB usuelle sur 8 bits (i50 ou p25) :

définition	1 image	débit	poids 1 min.	poids 1 h
SD 768x576	1.3 Mo	270 Mbps	2 Go.	120 Go
HD 1920x1080	6.2 Mo	1.2 Gbps	9 Go	540 Go
4K 3840x2160	24 Mo	4.8 Gbps	36 Go	2.2 To
8K 7680x4320	100Mo	17 Gbps	150 Go	8.8 To

Quantification RGB 10 bits avec fréquence élevée (p50) :

définition	1 image	débit	poids 1 min.	poids 1 h
SD 768x576	1.6 Mo	660 Mbps	5 Go	300 Go
HD 1920x1080	7.8 Mo	3.1 Gbps	23 Go	1.4 To
4K 3840x2160	31 Mo	12 Gbps	92 Go	5.6 To
8K 7680x4320	124 Mo	48 Gbps	370 Go	22 To

Pixels rectangulaire



Réduction de la définition en modifiant la forme des pixels :

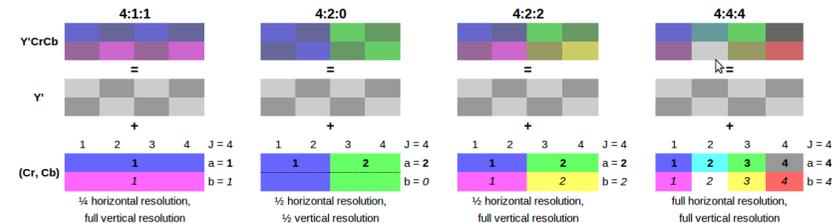
- passage du pixel carré au rectangulaire
 - 768 * 576 \rightarrow 704 * 576 (r1.09)
 - 1920 * 1080 \rightarrow 1440 * 1080 (r1.33)

Réduction du débit

Exemple de réduction du débit d'un flux vidéo HD :

pixel	composantes	comp. spatiale	comp. temp.	débit
carré	RGB (4:4:4)	1/1	1/1	1.2 Gbps
rect.	RGB (4:4:4)	1/1	1/1	950 Mbps
rect.	YCbCr (4:2:0)	1/1	1/1	480 Mbps
rect.	YCbCr (4:2:0)	8/1	1/1	60 Mbps
rect.	YCbCr (4:2:0)	8/1	5/1	12 Mbps

Sous-échantillonnage de la chrominance



- 4:4:4 YCbCr/RGB (absence de sous-échantillonnage)
- 4:2:2 (réduction horizontale de 1/2 de la taille des blocs Cb/Cr)
- 4:2:0 (réduction horizontale et verticale)

Compressions

Compressions spatiale et temporelles avec pertes ➡ voir plus loin...



Cinéma numérique

Recommandation DCI (*Digital Cinema Initiative*)

- 4:4:4 sur 12 bits
- colorimétrie X'Y'Z'
- définition 2K ou 4K
 - 2K r.2.39 (2048*858) ➡ 1,5 Gbps
 - 2K r.1.85 (1998*1080) ➡ 1,8 Gbps
 - 4K r.2.39 (4096*1716) ➡ 6 Gbps
 - 4K r.1.85 (3996*2160) ➡ 7,5 Gbps
- 16 canaux audio PCM 24 bits à 48/96 kHz
- **compression JPEG2000**
 - **2K : ~ 100 Mbps en VBR (max 250)**
 - **4K : ~ 200 Mbps en VBR (max 500)**
- d'autre fréquences que le 24p on été définies...



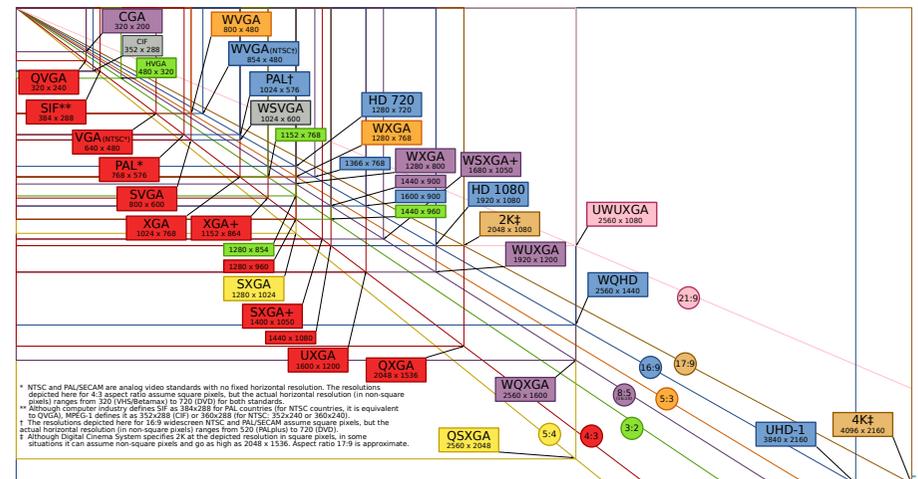
Cinéma analogique

Pellicule argentique :

- tri-acétate de cellulose de 35 mm de large,
- 4 perforations par images,
- défilement vertical à 24 images/s
- image de 24 mm * 16 mm (rapport 1.33)
- autres formats
 - 24 mm * 13 mm (rapport 1.85)
 - 24 mm * 10 mm (rapport 2.39)
 - 21 mm * 15 mm (rapport 1.37 avec audio)
 - 21 mm * 11 mm (rapport 1.85 avec audio)
 - 21 mm * 17 mm (anamorphique, rapport 2,39 et pistes audio)
- pouvoir résolvant du film : plus de 100 lignes/mm



Autres formats



CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

3 Contenu vidéo

vidéo numérique
encodages sans pertes
compressions avec pertes

SDI



Serial digital interface (SDI)

- normalisation en 1989 (SMPTE259M)
 - un coaxial de 75Ω
 - connecteur BNC
 - distance de 300 m
 - même câble que la vidéo analogique
- signal TVSD non compressé et non encrypté
 - vidéo : jusqu'à 10 bits, échantillonnage couleur 4:2:2
 - audio intégré : 16 canaux PCM 24 bits à 48 kHz
 - débit max : 270 Mbps

ITU-R BT.656

Normalisation du flux vidéo numérique de la TVSD

- TVSD (télévision basse définition)
 - 525 ligne ou 625 ligne en standard
 - flux entrelacé (envoi des 2 trames successivement)
 - Y, Cr, Cb en 4:2:2
 - échantillonnage de la luminance (Y)
8 bits sur 220 niveaux (valeurs de 16 à 235) ou 10 bits
 - échantillonnage de la chrominance (Cr, Cb)
codage de la différence de couleur sur 8 bits ou 10 bits
- débit brut de 270 Mbps (10 bits)
- profil RTP : **BT656** (RFC 2431)

Normes SMPTE

(Society of Motion Picture and Television Engineers)

Définis les évolutions de l'interface SDI :

Norme	Nom	Date	Débit	Application
SMPTE 259M	SD-SDI	1989	270 Mbps	480i, 576i
SMPTE 344M	ED-SDI	1995	540 Mbps	480p, 576p
SMPTE 292M	HD-SDI	1998	1.485 Gbps	720p, 1080i
SMPTE 424M	3G-SDI	2006	2.970 Gbps	1080p60
SMPTE ST-2081*	6G-SDI	2015	6 Gbps	4Kp30
SMPTE ST-2082*	12G-SDI	2015	12 Gbps	4Kp60, 8Kp30
SMPTE ST-2082*	24G-SDI	?	24 Gbps	4Kp120, 8Kp60

Les liens peuvent être doublé (dual link) ou quadruplés (quad link)

vidéo : jusqu'à 12 bits en 4:4:4 RGB

4K 60fps via Quad 3G-SDI



Codages vidéo avec perte

Vu les quantités d'informations, la plupart des codages vidéo effectuent une compression avec perte :

- ISO/IEC
 - M-JPEG, M-JPEG2000
 - MPEG-1, MPEG-2 Part 2, MPEG-4 Part 2/ASP, MPEG-4 Part 10/AVC et MPEG-H Part 2/HEVC
- ITU-T
 - H.261, H.262, H.263, H.264 et H.265
- Autres organisations/entreprises
 - VP3/Theora, VP6, VP7, VP8, VP9 et AV1
 - VC1, Dirac (VC2), VC3
 - Daala...

CONT : plan du cours 2/5

1 Introduction

contenus numériques
encodage des textes
encodage des images

2 Contenus audio

audio numérique
encodages sans pertes
compressions avec pertes

3 Contenu vidéo

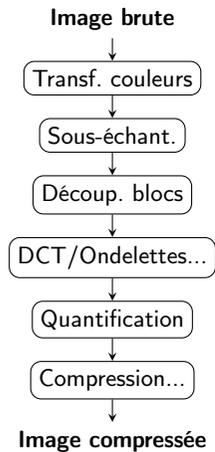
vidéo numérique
encodages sans pertes
compressions avec pertes

Réduction du débit : compression spatiale

Retour sur l'exemple de réduction du débit en HD :

pixel	composantes	comp. spatiale	comp. temp.	débit
carré	RGB (4:4:4)	1/1	1/1	1.2 Gbps
rect.	RGB (4:4:4)	1/1	1/1	950 Mbps
rect.	YCbCr (4:2:0)	1/1	1/1	480 Mbps
rect.	YCbCr (4:2:0)	8/1	1/1	60 Mbps
rect.	YCbCr (4:2:0)	8/1	5/1	12 Mbps

Compression spatiale : quels codages

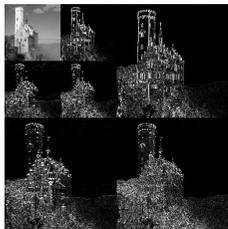
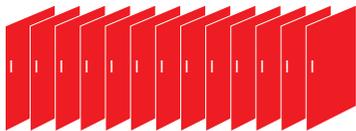


Les vidéo sont composées de suite d'image ➡ application des même techniques de compression

- approche sans perte insuffisante
- approche avec perte :
 - DCT (JPEG)
 - codages prédictifs
 - ondelettes (JPEG2000)
 - ...



M-JPEG2000



(CC BY-SA 3.0 Alessio Damato)

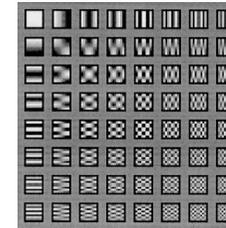
Motion JPEG2000
(Norme ISO/CEI 15444-3 :2002) :

- amélioration du format M-JPEG
 - transformée en ondelettes
 - codage arithmétique adaptatif
 - taux de compression : entre 1/10 et 1/50
 - absence de macrobloc
- existe en version sans-perte
- intégration possible de l'audio
- usage :
 - cinéma numérique (norme DCI)
 - indexation et montage vidéo
 - archivage
 - imagerie médicale
 - imagerie spatiale



M-JPEG

Motion JPEG :



- standard de fait
- chaque image (ou trame) est compressée indépendamment comme une image JPEG
- **pas de codage InterTrame**
- taux de compression limité : entre 1/5 et 1/20
- qualité dépendante du contenu vidéo
- complexité faible
- usage :
 - captation (caméra)
 - édition et montage vidéo



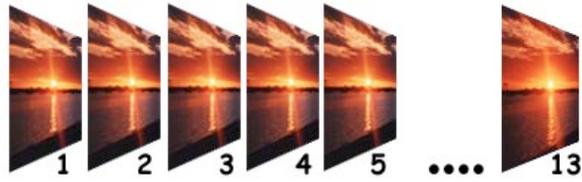
Réduction du débit : compression spatiale

Retour sur l'exemple de réduction du débit en HD :

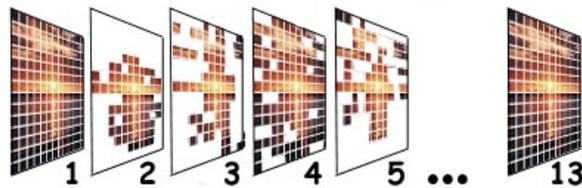
pixel	composantes	comp. spatiale	comp. temp.	débit
carré	RGB (4:4:4)	1/1	1/1	1.2 Gbps
rect.	RGB (4:4:4)	1/1	1/1	950 Mbps
rect.	YCbCr (4:2:0)	1/1	1/1	480 Mbps
rect.	YCbCr (4:2:0)	8/1	1/1	60 Mbps
rect.	YCbCr (4:2:0)	8/1	5/1	12 Mbps



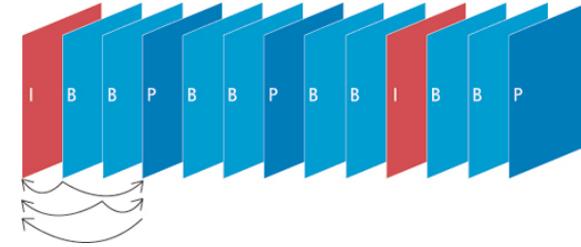
Redondance temporelle



- peu de différences entre 2 images successives :
 - blocs successifs identiques
 - blocs existants dans d'autres images
 - blocs nouveaux

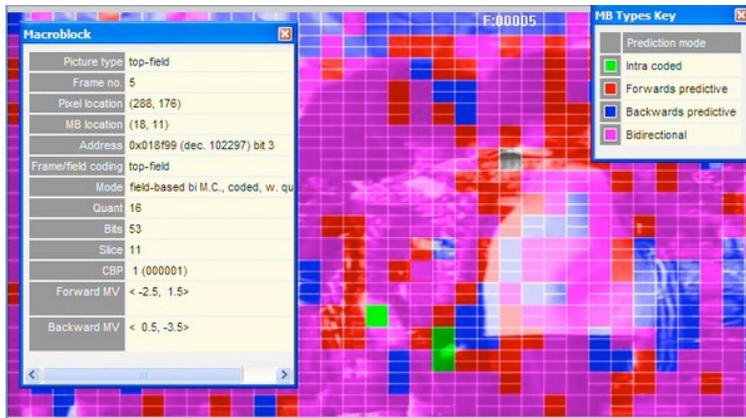


GOP : Group Of Pictures



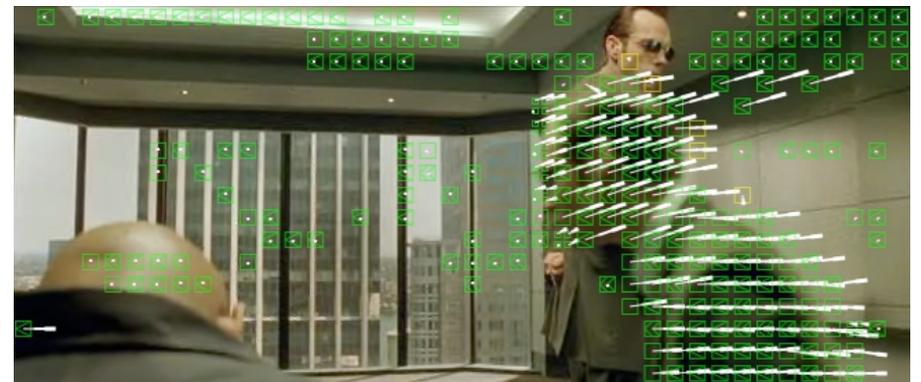
- analyse de l'image dans un groupe d'image limité
 - généralement 12 (Europe) ou 15 (USA) images → 0.5 s
- définition de 3 types d'image qui compose le GOP
 - **I** (Intra) image indépendante codée intégralement (blocs 100% intra)
 - **P** (Prédite) image composée à partir de blocs en amont (50% intra)
 - **B** (Bidirectionnelle) image construite à partir des images I et B autour d'elle (15% intra)
- GOP typique : **IBBPBBPBBPBI** (*Long GOP*)

Blocs / macroblocs



(Doc. Tektronix)

Prédiction du mouvement



Vecteurs de mouvements

- indication de l'origine antérieure d'un bloc

Compensation du mouvement

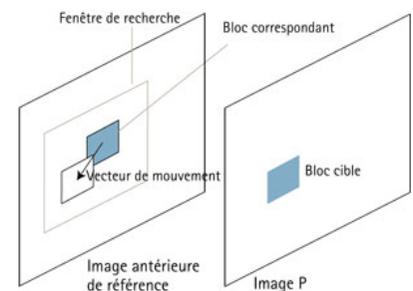
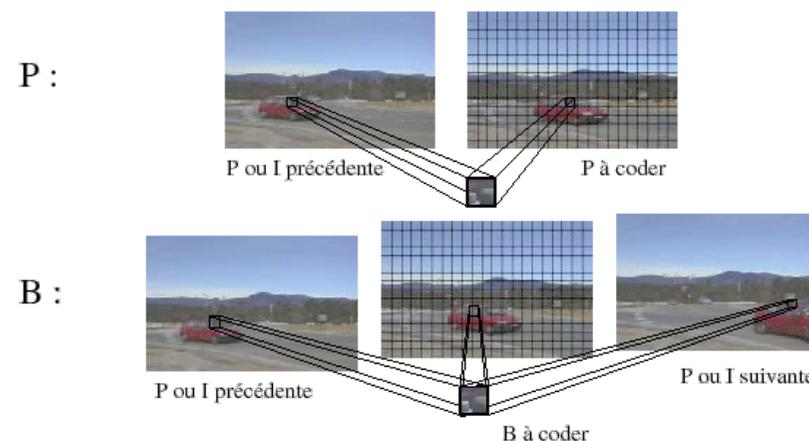


Illustration de la compensation de mouvement basée sur bloc.

- reconstruction des images **P**, pour chaque bloc :
 - prédiction arrière (*backward*) à partir du I ou P précédent
 - voisinage de 16 pixels dans toutes les directions (MPEG-2)
- reconstruction des images **B**
 - deux prédictions à partir des I ou P précédent et suivant (*forward*)
 - voisinages de 7 pixels dans toutes les directions (MPEG-2)
- calcul d'un bloc d'erreur associé (fortement compressé)

Compensation du mouvement



(Thèse Lionel Brunel)

MPEG-1

Norme ISO/CEI 11172 (1992)

Compression spatiale de MJPEG en ajoutant la **compression temporelle** :

- format typique proposé :
 - CIF (1/4 SDTV) :
 - 352 * 288 ou 352 * 240 en pixels rectangulaires
 - 384 * 288 ou 384 * 240 en pixels carré
 - 25 ou 30 images par seconde (sans entrelacement)
 - couleur 4:2:0
 - audio niveau 1, 2 et 3 (MP3)
- performances :
 - macrobloc 16x16
 - taux de compression de 1/20 à 1/30
 - débit vidéo de sortie de 1.2Mbps
- applications :
 - VCD (1.5 Mbps)
 - qualité type VHS

Conteneur MPEG1

Multiplexage audio/vidéo (.mpg) :

- codec vidéo (.mpv)
 - MPEG-1
- codec audio (.mpa)
 - MPEG layer 1
 - MPEG layer 2
 - MPEG layer 3
- exemple de fichier MPEG-1
 - débit global 1413 Kbps
 - vidéo MPEG-1 : 1150 Kbps, CBR, 352*288, aspect 4/3, 25 ips, progressif
 - audio MPEG layer 2, 254 Kbps, CBR, 2 canaux, 16 bits à 44.1 kHz

MPEG-2

Norme ISO/CEI 13818-2 (1994)

Amélioration du MPEG-1 pour la SDTV et la HDTV :

- format proposé
 - SDTV + 4*SDTV (TVHD) :
 - 704 * 576 ou 1440 * 1080 en pixels rectangulaires
 - 720 * 576 ou 1920 * 1080 en pixels carré
 - 50, 60, 100, 120 images par seconde (avec entrelacement)
 - audio : ajout du AAC (jusqu'à 48 canaux à 96 k)
- performances :
 - taux de compression de 1/30 à 1/40
 - débit vidéo de sortie de 6Mbps (SDTV) et 15 à 20Mbps (HDTV)
- applications :
 - DVD (9 Mbps)
 - TV numérique par satellite, par câble, ADSL et TNT

Profils d'encodage

Exemple : MP@ML (DVD), MP@HL1440 et MP@HL (BR)...

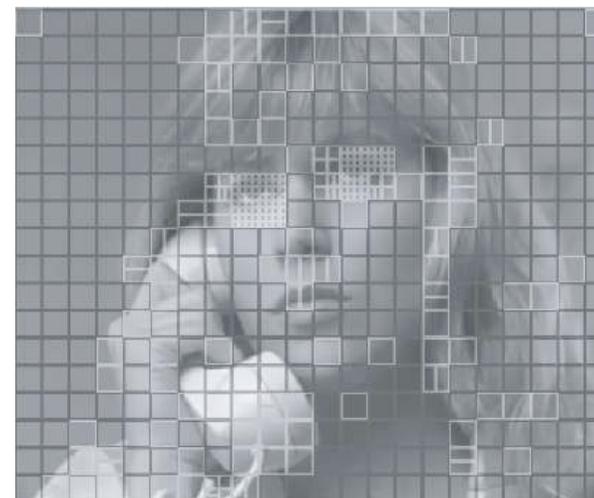
- 4:4:4
 - Level High (HL) = 4:4:4@1920x1080
 - Level High-1440 (HL-1440) = 4:4:4@1440x1080
 - Level Main (ML) = 4:4:4@720x576
- 4:2:2
 - Level High (HL) = 4:2:2@1920x1080
 - Level High-1440 (HL-1440) = 4:2:2@1440x1080
 - Level Main (ML) = 4:2:2@720x576
- Main(MP)
 - Level High (HL) = 4:2:0@1920x1080
 - Level High-1440 (HL-1440) = 4:2:0@1440x1080
 - Level Main (ML) = 4:2:0@720x576
 - Level Low (LL) = 4:2:0@352x288

conteneur MPEG-2

Multiplexage audio/vidéo (.mp2) :

- codec vidéo (.m2v)
 - MPEG-2/H262
- codec audio (.m2a)
 - MPEG layer 1 étendu
 - MPEG layer 2 étendu
 - MPEG layer 3 étendu
 - MPEG-2 AAC
- exemple de fichier MPEG-2
 - débit global 8520 Kbps
 - vidéo MPEG-2 : MP@ML, 9500 Kbps, VBR, 720*576, aspect 16/9, 25 ips, entrelacé
 - audio MPEG layer 2, 224 Kbps, CBR, 2 canaux, 16 bits à 48 kHz

Améliorer la compensation du mouvement



Utilisation de sous-blocs : 16x16, 8x16, 16x8, 8x8, 8x4, 4x8, et 4x4

GOP variable



(Images Jean-Charles Fouché)

Nouveaux profils d'encodage

Matrice liées à la complexité du décodage et au débit :

- Profiles
 - BP - Baseline (visio à très faible débit)
 - EP - Extended (mobile)
 - MP - Broadcast SD
 - HP (Fidelity Range Extensions) - Broadcast HD
 - H10P - 10 bits
 - H422P - ratio couleur 4:2:2
 - H444P - ratio couleur 4:4:4
- Levels (définition / débit / images par seconde)
 - 1 : 128x96@25 ou 176x144@15
 - ...
 - 5.1 : 4096x2304@25

MPEG-4

Amélioration du MPEG-2 (vidéo et objets interactifs) :

- deux codecs image
 - MPEG-4 ASP Part 2
 - MPEG-4 AVC Part 10 (H264)
 - prise en charge de nombreux formats (de la visio à la 4K)
 - audio : AAC (jusqu'à 48 canaux à 96 kHz)
- performances :
 - blocs de taille variable
 - GOP de taille variable
 - formats multiflux
 - taux de compression de l'ordre de 1/50
 - débit vidéo de sortie de 1,5 à 7 Mbps (SDTV) et de 7 à 14 Mbps (HDTV)
- applications :
 - BR (30 Mbps)
 - TV numérique de seconde génération par satellite, par câble ADSL et TNT

conteneur MPEG-4

Multiplexage audio/vidéo (.mp4) :

- codec vidéo (.m4v)
 - MPEG-4 SP
 - MPEG-4 ASP
 - MPEG-4 AVC/H264
 - H263
- codec audio (.m4a et .m4p)
 - MPEG layer 3 étendu
 - MPEG-4 AAC
 - MPEG-4 AMR, ALS, CELP et TwinVQ
- exemple de fichier MPEG-4
 - débit global moyen 1520 Kbps (Max 2420, Min 1100)
 - vidéo AVC/H264 : HP@3.0, 9500 Kbps, VBR, 480*576 anamorphosé, aspect 4/3, 25 ips, entrelacé
 - audio AAC, 192 Kbps, CBR, 2 canaux, 16 bits à 48 kHz

Autre codec de même génération que MPEG-4

Theora (Xiph.org) : libre issu de VP3

Dirac (BBC) : libre

VC1 (Microsoft) : propriétaire (utilisé également dans les DVD)

TrueMotion (On2 Technologies)

- VP5, VP6, VP7 : propriétaires utilisés dans Flash, Skype
- VP8 : libre (On2 racheté par Google)

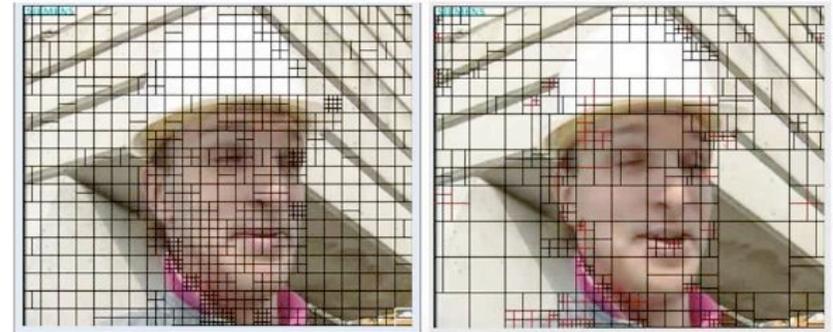
➡ **WebM (alternative libre pour la vidéo HTML5) utilise VP8 + Vorbis**

MPEG HEVC / H265

High Efficiency Video Coding

- nouveaux profils
 - Main 12, Main 4:2:2 10, Main 4:2:2 12, Main 4:4:4, Main 4:4:4 10, Main 4:4:4 12, Monochrome 12, Monochrome 16, Main 12 Intra, Main 4:2:2 10 Intra, Main 4:2:2 12 Intra, Main 4:4:4 Intra, Main 4:4:4 10 Intra, Main 4:4:4 12 Intra, Monochrome 12 Intra et Monochrome 16 Intra
 - prise en charge de nombreux formats (de la visio à la TVUHD 8K)
 - 100, 120 et 150 images par secondes
- performances :
 - blocs de taille variable
 - GOP de taille variable
 - taux de compression de l'ordre de 1/100

Améliorer encore plus la compensation du mouvement



Utilisation de sous blocs : découpages arborescent (*quadtree*) à partir d'un bloc 64x64

Autre codec de même génération que H265/MPEG HEVC

AV1 (VP9 de Google + Thor de Cisco +...) : libre

Dalaa (Xiph.org) : libre

➡ **WebM (alternative libre pour la vidéo HTML5) évolue en 2013 vers VP9 + Opus**

➡ **IETF NETVC initiative : AV1 et Dalaa sont candidats**