

Sistemas de Computação

Mestrado Integrado em Engenharia de Comunicações

2012/2013

Dados (I)

- Computadores
 - Processam e armazenam todos os tipos de dados em binário;
- Comunicação entre pessoas
 - Linguagem, imagens, sons;
- Formatos dos dados
 - Existem especificações para converter os dados em algo usável pelo computador;
 - Define as diferentes formas em que os dados usados pelas pessoas pode ser representados, guardados e processados num computador.
 - Existem variados dispositivos de entrada (*input*) que podem ser usados. A escolha dos dispositivos depende do formato original dos dados e da representação dos dados que se quer ter no computador.
 - Exemplo: teclado é algo simples e directo – cada tecla gera um código binário que é identificado como uma representação de um dado carácter
 - Um som já é mais complicado de representar, dado que é algo que muda de forma contínua ao longo do tempo.

Dados (II)

- Binário
 - É de natureza discreta;
 - Exemplo: teclado, em que o utilizador carregar no “A”
 - O teclado gera um código binário para cada tecla
- Analógico
 - Dados em contínuo, como o som e as imagens
 - É preciso hardware próprio para converter os dados em números binários;

Dados (III)

- Metadados
 - Informação que descreve ou permite interpretar o significado dos dados
 - Por exemplo, para representar uma imagem é preciso saber:
 - Número de cores representado em cada ponto da imagem;
 - O método usado para representar a cor;
 - O número de pontos que constituem a imagem em termos de largura e altura;
- Cada programa pode guardar e processar os dados no formato que desejar
 - Exemplo: Word e WordPerfect são dois processadores de texto que guardam os dados de forma diferente;
 - Os formatos usados por cada programa são chamados de **formatos proprietários**;
- Representações normalizadas (*standard*) de dados
 - Permite a interação:
 - Diferentes programas;
 - Entre os programas e os dispositivos de I/O;
 - Entre diferentes peças de hardware;
 - Entre sistemas que partilham dados através das redes ou de dispositivos transportáveis como CD-ROMs
 - Os dados tem de ser reconhecidos por uma grande variedade de hardware e de software para permitir que possam ser usados por utilizadores com diferentes sistemas computacionais

Dados (IV)

- Representação de dados usuais

Tipo de dados	Standards
Alfa numéricos	Unicode, ASCII, EBCDIC
Imagem (bitmap)	GIF – Graphical Image Format TIFF – Tagged image file format PNG – Portable network graphics JPEG – Joint Photographic Experts Group
Imagem (objecto)	PostScript SWF – Macromedia Flash SVG – Scalable Vector Graphics
Fontes	TrueType
Som	WAV – Waveform Audio File Format AVI – Audio Video Interleave MP3 – MPEG-1 Audio Layer-3 MIDI – Musical Instrument Digital Interface WMA – Windows Media Audio

Tipo de dados	Standards
Descrição de páginas	PDF – Adobe Portable Document Format HTML – HyperText Markup Language XML – Extensible Markup Language
Video	Quicktime MPEG-2 ou MPEG-4 RealVideo WMV - Windows Media Video DivX - Digital Video Express

Representação interna de dados

- Reflecte:
 - Complexidade da fonte
 - Tipos de processamento necessário
- Compromissos:
 - Precisão e resolução
 - Simples foto vs pintura num livro de arte
 - Densidade (arquivo e transmissão)
 - É necessária uma maior quantidade de dados para aumentar a precisão e resolução
 - Compressão permite representar os dados de forma mais compacta
 - Metadados: dados que descrevem ou interpretam o significado dos dados
 - Facilidade da manipulação
 - Processar um áudio simples vs um som de alta fidelidade
 - Normalização
 - Formatos proprietários para arquivar e processar os dados (exemplo: word)
 - Normas / standards de facto: normas proprietárias baseadas em sistemas aceites de forma generalizada (exemplo: PostScript)

Tipos de dados: alfa-numéricos

- Alfa-numéricos
 - Caracteres b H p
 - Dígitos numéricos: 7 9 3
 - Sinais de pontuação: ! , ?
 - Caracteres com significado especial: \$ £ &
- Caracteres numéricos vs números
 - Ambos são introduzidos como caracteres normais
 - O computador “converte” para números para efeitos de cálculos
 - Tratados como caracteres se forem processados como texto
 - Exemplo: número de telefone, código-postal

Códigos alfa-numéricos

- Conjunto de bits representam um carácter;
- Valor do número binário que representa o carácter respeita a ordem no alfabeto
 - Facilita a ordenação e a pesquisa

Representação de caracteres

- ASCII - esquema de codificação mais utilizado
- EBCDIC – Sistemas IBM (legado)
- Unicode – Desenvolvido para utilização universal

ASCII (I)

- Desenvolvido pela ANSI – *American National Standards Institute*
- Representa
 - Alfabeto latino
 - Numeração árabe
 - Caracteres de pontuação *standard*
 - Pequeno conjunto de acentos e outros caracteres usados em línguas europeias
- Código de 7 bits: 128 caracteres

ASCII (II)

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

ASCII (III)

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	@	96	60	`
1	01	Start of heading	33	21	!	65	41	A	97	61	a
2	02	Start of text	34	22	"	66	42	B	98	62	b
3	03	End of text	35	23	#	67	43	C	99	63	c
4	04	End of transmit	36	24	\$	68	44	D	100	64	d
5	05	Enquiry	37	25	%	69	45	E	101	65	e
6	06	Acknowledge	38	26	&	70	46	F	102	66	f
7	07	Audible bell	39	27	'	71	47	G	103	67	g
8	08	Backspace	40	28	(72	48	H	104	68	h
9	09	Horizontal tab	41	29)	73	49	I	105	69	i
10	0A	Line feed	42	2A	*	74	4A	J	106	6A	j
11	0B	Vertical tab	43	2B	+	75	4B	K	107	6B	k
12	0C	Form feed	44	2C	,	76	4C	L	108	6C	l
13	0D	Carriage return	45	2D	-	77	4D	M	109	6D	m
14	0E	Shift out	46	2E	.	78	4E	N	110	6E	n
15	0F	Shift in	47	2F	/	79	4F	O	111	6F	o
16	10	Data link escape	48	30	0	80	50	P	112	70	p
17	11	Device control 1	49	31	1	81	51	Q	113	71	q
18	12	Device control 2	50	32	2	82	52	R	114	72	r
19	13	Device control 3	51	33	3	83	53	S	115	73	s
20	14	Device control 4	52	34	4	84	54	T	116	74	t
21	15	Neg. acknowledge	53	35	5	85	55	U	117	75	u
22	16	Synchronous idle	54	36	6	86	56	V	118	76	v
23	17	End trans. block	55	37	7	87	57	W	119	77	w
24	18	Cancel	56	38	8	88	58	X	120	78	x
25	19	End of medium	57	39	9	89	59	Y	121	79	y
26	1A	Substitution	58	3A	:	90	5A	Z	122	7A	z
27	1B	Escape	59	3B	;	91	5B	[123	7B	{
28	1C	File separator	60	3C	<	92	5C	\	124	7C	
29	1D	Group separator	61	3D	=	93	5D]	125	7D	}
30	1E	Record separator	62	3E	>	94	5E	^	126	7E	~
31	1F	Unit separator	63	3F	?	95	5F	_	127	7F	□

ASCII (III)

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
128	80	Ç	160	A0	á	192	C0	Ł	224	E0	α
129	81	ü	161	A1	í	193	C1	ł	225	E1	β
130	82	é	162	A2	ó	194	C2	Ť	226	E2	Γ
131	83	â	163	A3	û	195	C3	ţ	227	E3	Π
132	84	ä	164	A4	ñ	196	C4	—	228	E4	Σ
133	85	à	165	A5	Ñ	197	C5	†	229	E5	σ
134	86	ã	166	A6	•	198	C6	‡	230	E6	μ
135	87	ç	167	A7	°	199	C7	‡	231	E7	τ
136	88	ê	168	A8	¿	200	C8	Ł	232	E8	φ
137	89	ë	169	A9	ŕ	201	C9	ŕ	233	E9	θ
138	8A	è	170	AA	ŕ	202	CA	Ł	234	EA	Ω
139	8B	ì	171	AB	•	203	CB	ŕ	235	EB	δ
140	8C	í	172	AC	•	204	CC	ŕ	236	EC	∞
141	8D	î	173	AD	•	205	CD	=	237	ED	ω
142	8E	Ā	174	AE	«	206	CE	ŕ	238	EE	ε
143	8F	Ā	175	AF	»	207	CF	Ł	239	EF	Π
144	90	Ē	176	B0	•	208	D0	Ł	240	FO	≡
145	91	æ	177	B1	•	209	D1	ŕ	241	F1	±
146	92	æ	178	B2	•	210	D2	ŕ	242	F2	≥
147	93	ó	179	B3		211	D3	Ł	243	F3	≤
148	94	ö	180	B4	†	212	D4	Ł	244	F4	[
149	95	ò	181	B5	†	213	D5	ŕ	245	F5]
150	96	û	182	B6	ŕ	214	D6	ŕ	246	F6	÷
151	97	ù	183	B7	ŕ	215	D7	ŕ	247	F7	≈
152	98	ÿ	184	B8	ŕ	216	D8	†	248	F8	"
153	99	Ö	185	B9	ŕ	217	D9	ŕ	249	F9	•
154	9A	Û	186	BA	ŕ	218	DA	ŕ	250	FA	·
155	9B	◊	187	BB	ŕ	219	DB	■	251	FB	√
156	9C	£	188	BC	ŕ	220	DC	■	252	FC	∞
157	9D	¥	189	BD	ŕ	221	DD	■	253	FD	≈
158	9E	₯	190	BE	ŕ	222	DE	■	254	FE	■
159	9F	ƒ	191	BF	ŕ	223	DF	■	255	FF	□

Fonte: <http://www.cdrummond.qc.ca/cegep/informat/professeurs/alain/files/ascii.htm>

EBCDIC (I)

- EBCDIC - *Extended Binary Coded Decimal Interchange Code*
 - Desenvolvido pela IBM
 - Usado essencialmente pela IBM e por alguns sistemas “compatíveis com IBM”;
 - É possível a conversão de/para ASCII
 - Códigos dos caracteres são diferentes no ASCII e no EBCDIC

EBCDIC (II)

		EBCDIC character codes															
		1st hex digit															
2nd hex digit		0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	DS		SP	&	-										0
1	SOH	DC1	SOS				/			a	j			A	J		1
2	STX	DC2	FS	SYN						b	k	s		B	K	S	2
3	ETX	TM								c	l	t		C	L	T	3
4	PF	RES	BYP	PN						d	m	u		D	M	U	4
5	HT	NL	LF	RS						e	n	v		E	N	V	5
6	LC	BS	ETB	UC						f	o	w		F	O	W	6
7	DEL	IL	ESC	EOT						g	p	x		G	P	X	7
8		CAN								h	q	y		H	Q	Y	8
9		EM								i	r	z		I	R	Z	9
A	SMM	CC	SM		C CENT	!	:										
B	VT	CU1	CU2	CU3		\$,	#									
C	FF	IFS		DC4	<	*	%	@									
D	CR	IGS	ENQ	NAK	()	_	'									
E	SO	IRS	ACK		+	:	>	=									
F	SI	IUS	BEL	SUB		--	?	"									

Fonte: <http://www.diycalculator.com/sp-ascii.shtml>

Unicode

- Usa 16 bits – representa 65536 caracteres
- A norma internacional ISO/IEC 10646 define o **Conjunto Universal de Caracteres**
 - Unicode Consortium trabalha em conjunto com a ISO na definição dos caracteres
- ASCII Latin-I é um subconjunto do Unicode
 - Corresponde aos caracteres 0-255 da tabela Unicode
- Suporta múltiplas línguas
 - Suporta quase os caracteres de todos os alfabetos
 - Suporta os símbolos usados nas línguas chinesas, japonês e coreano
 - Suporta caracteres “compostos” usados em algumas línguas

Códigos de caracteres

- A ordem dos caracteres na tabela ASCII é importante
 - Permite ordenar com facilidade
- Classes de códigos:
 - Caracteres imprimíveis
 - Aparecem visíveis no ecrã e na impressora
 - Caracteres de controlo
 - Controlo da posição no ecrã ou na impressora
 - VT: Vertical Tab
 - LF: Line Feed
 - FF: Form Feed
 - Causam uma acção
 - BEL: campainha
 - DEL: apaga o carácter anterior
 - Comunicação de estado entre o computador e o dispositivo de I/O
 - ESC: altera o significado dos caracteres seguintes

Teclado

- Scan code
 - 2 *scan code* no teclado
 - Um é gerado quando a tecla é pressionada e o outro quando se solta a tecla
 - Convertido para Unicode, ASCII, EBCDIC por software incluído no PC
- Vantagem
 - Fácil de adaptar a diferentes linguagens e diferentes *layout* de teclados;
 - Os *scan codes* separados permitem múltiplas combinações de teclas
 - Exemplo: teclas de *shift* e de *control*

Outras formas de leitura de alfa-numéricos

- OCR – Optical Character Reader/Recognition
 - Leitura (scan) do texto e introduz o texto (caracteres)
- Leitores de códigos de barras
 - Usados em aplicações que requerem a introdução rápida, sem erros ou repetitiva de dados, com um treino mínimo dos funcionários
 - Exemplos: supermercados, controlo de inventário
- Leitores de banda magnética
 - Leitura de dados alfa-numéricos de cartões (de crédito/débito)
- RFID – Radio Frequency Identification
 - Leitura e transmissão de dados entre etiquetas RFID e computadores
- Voz
 - É fácil grava a voz num formato digital mas difícil de converter para alfa-numérico
 - Requer conhecimentos de padrões de som em linguagem, regras de pronunciação, gramática e sintaxe

Imagens

- Fotografias, figuras, ícones, desenhos, gráficos
- Duas opções:
 - *Bitmap* ou *raster images*
 - *Object* ou *Vector images*
 - Composto de objectos gráficos como linhas e curvas definidas geometricamente
- Diferenças:
 - Qualidade da imagem;
 - Espaço de armazenamento necessário;
 - Tempo de transmissão;
 - Facilidade de modificação

Imagens bitmap

- Usado para imagens realísticas, com uma variação contínua d sombra, cor, formato e textura
 - Exemplos:
 - Fotografias digitalizadas
 - Imagens geradas pelo *paint* do *windows*
- Preferível quando a imagem contém grande quantidade de detalhe e os requisitos de processamento são simples
- Input:
 - Scanners;
 - Máquinas fotográficas digitais;
 - Criado no PC
- Gestão feita por software de edição de fotografias e similares

Imagens bitmap

- Cada pixel individual é um elemento gráfico guardado com um número em binário;
 - Pixel: pequeno ponto associada a uma dada coordenada
- Monocromática
 - 1 bit por pixel
- Escala de cinzento: preto, branco ou 254 níveis de cinzento
 - 1 byte por pixel
- Cores: 16 cores, 256 cores, 24-bit *true color* (16.7 milhões de cores)

Imagens bitmap



1 bit p&b



8 bits
cinzento

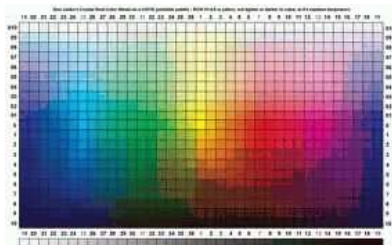


8 bits

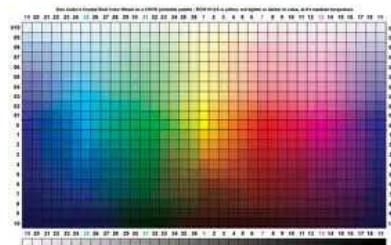


24 bits

Imagens bitmap



8 bits



24 bits

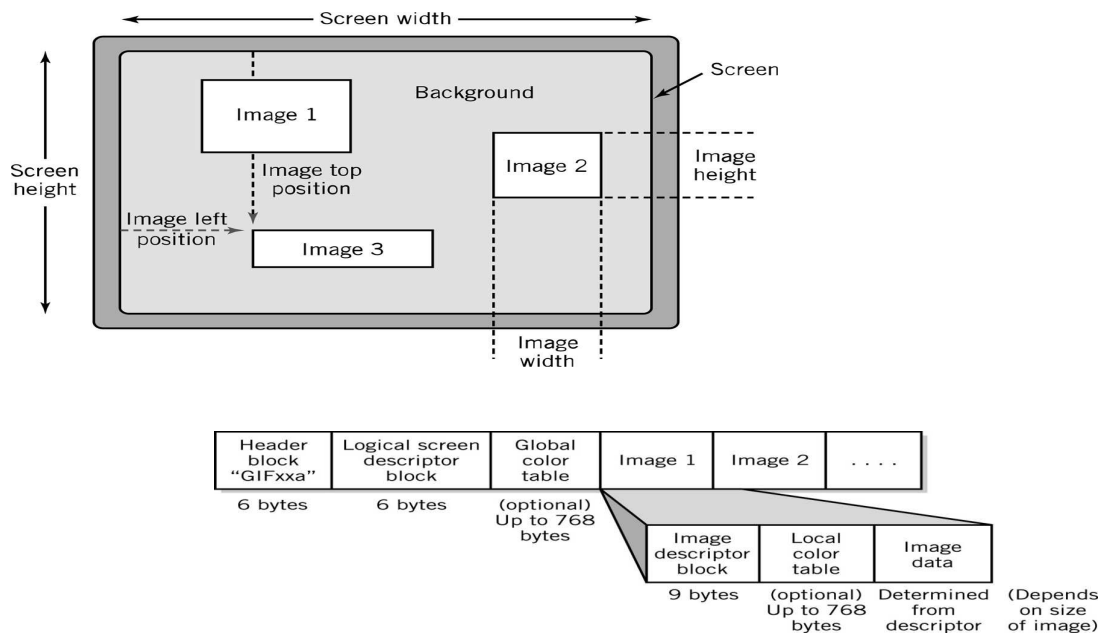
Imagens bitmap

- Ficheiros de grande dimensão
 - Exemplo: 600 linhas por 800 pixel com 1 byte por cada uma das três cores → ~1,5 Mb
 - $600 \times 800 = 480.000 \text{ pixels} \times 3 \text{ bytes} = 1.440.000$
- Tamanho do ficheiro determinado por:
 - Resolução (número de pixels por polegada)
 - Nível de detalhe que afecta a clareza o *sharpness* de uma imagem
 - *Levels*: número de bits usados para definir um nível de cinzento ou número de cores
 - Paleta: tabela de cores onde se usa um código para cada pixel em do valor da cor
 - Compressão dos dados

GIF – Graphics Interchange Format

- Desenvolvido pela CompuServe em 1987
- GIF89a permite construir imagens animadas
 - Permite que as imagens sejam mostradas em sequência, com intervalo de tempo definido
- Limite no número de cores: 256
- Compressão da imagem usando o algoritmo LZW – Lempel-Fif-Welch
- Usado em imagens que sejam desenhos de linhas, clip art e outras onde surgem grandes blocos da mesma cor sólida
- Compressão sem perda de dados (*lossless compression*)

GIF – Graphics Interchange Format



Fonte: Englander, Irv; "The architecture of Computer Hardware, system software & networking"

JPEG

- JPEG – *Joint Photographers Expert Group*
- Permite mais de 16 milhões de cores
- Popular na área da fotografia
- Usa um algoritmo de compressão que perde informação (*lossy compression*)
 - Perde alguns dados de forma a diminuir o tamanho do ficheiro e permitir a sua transmissão de forma mais célere
 - Pode levar à redução da qualidade das imagens, tendendo a distorcer as linhas rectas (*sharp lines*)

Imagem como objectos (*object images*)

- Imagens criadas por ferramentas de desenho (Exemplo: Illustrator) ou como resultado de aplicações específicas (ex.: gráficos criados numa folha de cálculo)
- Compostas por linhas e formas de várias cores
- O computador traduz formulas geométricas para criar gráficos
- O espaço ocupado pelo ficheiro depende da complexidade da imagem
 - Normalmente gera ficheiros mais pequenos do que os bitmaps
- Baseada em formulas matemáticas
 - Fácil de mover, de dimensionar e rodar sem perder qualidade/forma (ao contrário do que acontece com os bitmap)
- Não pode ser usado em fotografias ou pinturas
- Não pode ser mostrado ou impresso de forma directa
 - Tem de se converter para bitmap, dado que os dispositivos de *output* mostram imagens bitmap (com excepção das plotters)
- Há filmes de cinema criados com este tipo de imagens

PostScript

- *Page Description Language*
 - Lista de procedimentos e instruções que descrevem dada um dos objectos a ser impressos numa página
 - Armazenado como um ficheiro de texto ASCII ou Unicode
 - Um programa de interpretação de postscript no dispositivo de *output* lê o conteúdo do ficheiro e gera a imagem correspondente
 - Inclui fontes que podem ser usadas, como qualquer outro objecto

Bitmap vs Object images

Bitmap (raster)	Object (vector)
Mapa de pixels	Formas definidas geometricamente
Qualidade fotográfica	Desenhos complexos
Software de tratamento de imagens	Software de desenho vectorial
Maiores necessidades em termos de espaço de armazenamento	Maiores necessidades em termos de processamento
Aumentar as imagens causa perda de qualidade	Objectos aumentam de forma contínua
Resolução de <i>output</i> limitada pela resolução da imagem	Resolução de <i>output</i> limitada pela resolução do dispositivo de saída

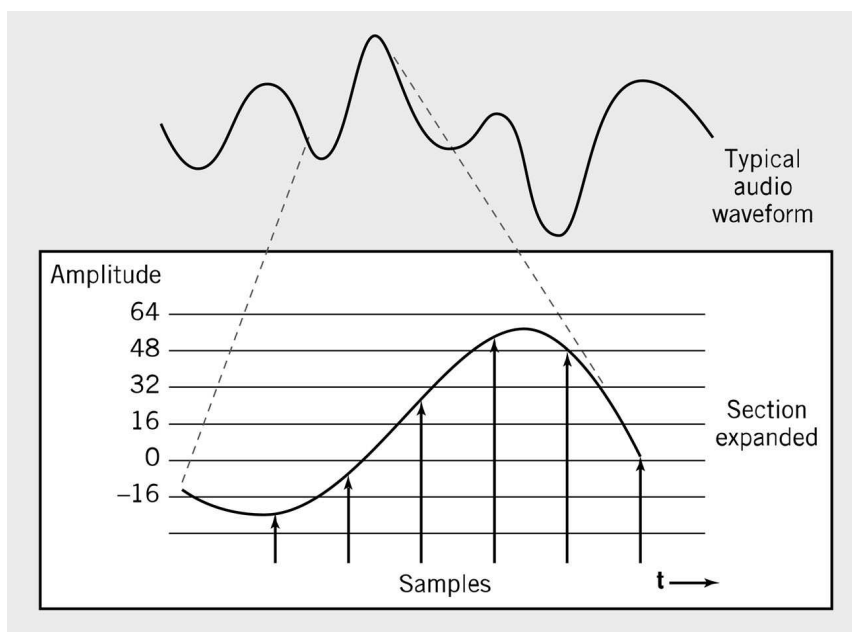
Vídeo

- Quantidades massivas de dados
 - Exemplo 1
 - Câmara de vídeo produz um vídeo com 640 x 480 pixels, em *true color*, gravando 30 *frames/segundo* -> 27,65Mb/Segundo
 - Filme com 1 minuto: 1,6Gb
 - Exemplo 2
 - $1024 \times 768 \text{ pixel} \times 3 \text{ bytes/cor} \times 30 \text{ frames/segundo} = 70,8 \text{ Mb/segundo}$
 - Filme de 1 minuto: 4,25 Gb
- Opções para reduzir o tamanho do ficheiro
 - Diminuir a resolução
 - Limitar o número de cores
 - Reduzir o *frame rate*
 - Comprimir o vídeo
- A opção depende do destino do vídeo
 - *Streaming* de vídeo: é enviado pela Internet e visualizado ao mesmo tempo que é recebido
 - Dados locais: como um DVD ou um ficheiro armazenado no computador

Áudio

- Requisitos para transmissão e processamento são inferiores aos que existiam no vídeo
- O som é analógico por natureza
 - Para o digitalizar recolhe-se amostras em intervalos regulares de tempo
 - Usa-se um conversor A-D
- A velocidade de recolha de amostras deve ser suficientemente elevada para captar todas as nuances do som
 - Número de vezes, por segundo, que o som é medido durante o processo de gravação
 - 1000 amostras por segundo = 1Khz
 - CD de música: 44,1 khz
 - Cada amostra pode ser gravada com 8 bits, 16 bits ou 2x16 bits no caso do som stereo

Áudio



Exemplos de formatos de áudio

- MP3
 - Deriva do MPEG-2 (ISO Moving Picture Experts Group)
 - Remove os sons que o ouvinte não consegue ouvir ou não vai notar que foram removidos
 - Usa várias técnicas para criar ficheiros “pequenos”
 - Valores típicos: 44.100 Hz, 128 ou 192kb/segundo
- WAV
 - Desenvolvido pela Microsoft como parte da sua especificação multimédia
 - Formato genérico, usado para armazenar e reproduzir pequenos sons

Compressão de dados (I)

- Compressão
 - Guardar os dados de forma a ocupar menos espaço
- Rácio de compressão
- *Lossless* – sem perdas
 - O algoritmo inverso permite obter os dados originais de forma exacta.
 - Exemplo: TIFF, GIF, PCX
- *Lossy* – Com perdas
 - Balanço entre os dados que se perde e o tamanho do ficheiro
 - Rácios de compressão mais elevados: chegando a 10:1
 - Usado habitualmente em multimédia
 - Exemplo: JPEG

Compressão de dados (II)

05573200001473291000006682732732...

0155732041473291056682732732...

732 -> Z

0155Z0314Z91056682ZZ...