

マルチメディアデータ


文字の符号化

- 文字コード：ビットパターンと「文字」を対応づけるルール
- 文字コードセット（以降，文字セット）
 - 数値と文字の対応
 - どこまでの「文字」を扱うか？
- 文字エンコーディング
 - 文字コードセットによって数値化された文字をどのようなビットパターンにするか

文字セット	ビット数	文字種	備考
ASCII	7bit	制御文字, 数字, アルファベット	
ISO-8859-1	8bit	ASCII + アクセント	Latin-1
JIS X 0201	8bit	ASCII + カタカナ	半角カタカナ
JIS X 0208	16bit	全角かな / カナ, 漢字	制御文字ははASCIIと同じ
Unicode	4byteまで	世界中の文字	16bitまでの範囲をBMP (Basic Multilingual Plane)

文字コード	文字セット	備考
ASCII	ASCII	そのまま使う
ISO-2022-JP	ASCII, JIS X 0208	文字セットを切り替える
Shift_JIS	JIS X 0201, JIS X 0208	ASCIIと互換性
EUC-JP	ASCII, JIS X 0208	ASCIIと互換性
UTF-8	Unicode	ASCIIと互換性, 1から4bytes可変長
UTF-16	Unicode	バイトオーダーによる違い

画像

- 画像オブジェクトの形式
 - ビットマップ（ラスタデータ）：「色」の配置
 - ドロー（ベクタデータ）：「点」と「線」の配置
- ラスタ形式
 - 格子上のすべての点に配置する「色」を決める（色情報を持った点  ピクセル）
 - 写真などの複雑なデータも表現可能
 - 拡大・縮小などの情報量の増減が不得意
- ベクタ形式
 - 格子上の「点」とそれをつなぐ「線」の特徴（曲がり具合や色）を決める
 - 写真などの複雑なデータは表現できない（イラストなどに適する）
 - 拡大・縮小などの情報量の増減が得意

色データ

- RGB（光による表現）：ディスプレイなど
 - R (Red), G (Green), B (Blue) のそれぞれの強さ（濃淡）で色を表現
 - RGBが全部混ざると白になる
- CMYK（絵の具による表現）：プリンターなど
 - C (Cyan), M (Magenta), Y (Yellow) のそれぞれの強さ（濃淡）で色を表現
 - CMYが全部混ざると黒になる
 - K (Key plate) は通常黒で CMY で作る黒を補強

音


- 空気の連続的かつ周期的な圧力変化（波の性質）
- 波の周期が速い → 高い音
- もとの波を復元できるようなデータを格納する
 - 圧力の大きさ

サンプリング

- 「大きさの変化」（つまり「速さ」）を記録する必要がある
- 連続する「瞬間」を切り取り「速さ」を記録
- どの「瞬間」を切り取ればいいのか？
 - ▶ 最大周波数の2倍より大きい周波数で観測することでもとの情報を復元できる

※周波数 = 1/周期

音データ

- 観測される数値（振幅）を量子化
- 量子化の方法で種類がある：線形  リニアPCM
- CD-DA（CD音源）：サンプリング周波数 44.1kHz，量子化ビット数 16bit

動画

- 映像データ（連続する画像データ）と音データ
- フレームレート
 - 1秒間は何画像で構成されるか
- 動画はデータ量が多くなるので、通常はデータ圧縮が行われる
 - 前の画像との差分だけを記録，など
 - 動画クオリティ
 - ビットレート：毎秒あたりどのくらいのビット数にするかの目安
 - YouTube 1080p（フルHD）：1,000 から 3,000 kbps