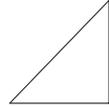


#2 bit,Byte, デジタル化の自由度

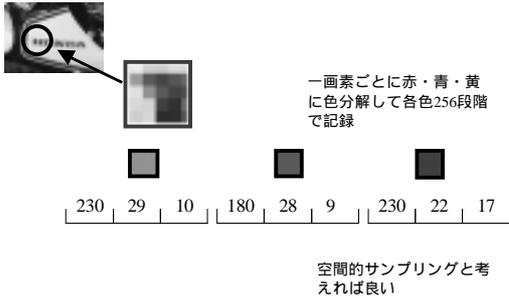
2003, Spring term, Yutaka Yasuda

デジタル処理の原理

- デジタル情報化 = 符号化
 - 対象の特徴を記号や数値によって確定的に表現
 - アナログ的表現
 - 三角形を真似て描く
 - デジタル的表現
 - 三角形の頂点の座標位置を(0,0),(100,0),(100,210)と記述
 - これがデジタルデータそのもの



画像データの例



符号化 (デジタルデータの表現)

- どんなものでも特徴を記号(数値)化できればデジタル情報に変えられる
- サンプリング(標本化)という手法
 - 音: 波を一定時間で区切って測定
 - 写真(静止画像): 一定間隔のマスに区切って色分解
 - テレビ(動画像): 時間と空間の両方で標本化
 - パラパラマンガのように映像を一定時間で区切って、連続した静止画として処理
 - それに音を加える
- 一定のルールで値を測定
 - この値がデジタルデータそのものになる

データ

- その実体は数値(記号)の列
 - 音声: 111,121,122,89,80,82,75....
 - 静止画: 10,240,22,30,34,80...
 - 音声付き動画: 12,33,45,1123,488...
- これだけでは無意味
 - 符号化ルールとデータは常に一体
 - それがどんなものか
 - どのようにして数値化したのか
 - どうすれば再現(利用可能な状態に)できるか
- このルールがフォーマットを生む

フォーマット(書式)

- デジタルデータを解釈するには
 - 解釈(解読)ルールが必要
 - データそのものはただの記号(数値)の列
- フォーマット(書式)
 - つまりデータにはフォーマットがある
 - フォーマットを間違えて解釈すると間違った結果が導き出される(いわゆる「文字化け」など)
 - 異なるアプリケーションでデータが扱えない理由
 - データにおける「互換性」という概念の実体

事例

- Microsoft の場合
 - フォーマットを閉ざして市場を囲い込み
 - アクセスするためには買わざるを得ない
- Adobe Acrobat の場合
 - フォーマットは殆ど公開しているようなもの
 - 読むだけのソフトを無料で積極的に配る
 - 作成ソフトウェアを売る
- Open Source への動き
 - アプリケーションは消えても良いがデータは残る
 - 永続的なアクセスを求めてオープンなものへと

デジタル化の特長

- 初期ノイズの発生
 - デジタル化(数値化)の時点でオリジナルとの相違が発生
 - デジタル化処理とは近似のことである
 - 量子化雑音と呼ぶ
- 伝達や記録、複製に伴うノイズの抑制
 - ゼロに出来るかも知れない
 - 完全に同一内容の複製の作成が可能
- 因みに、、、
 - 伝達・記録・複製はどれも同じこと
 - 技術にその境界はない

デジタル化の特長

- 数値処理のテクニック(数学的手法)が使える
 - 通信におけるエラーチェック(検証)、圧縮など
- 多様な資源・技術を利用できる
 - データの表現が汎用である(符号であれば良い)
 - 符号が乗るメディア・通信路なら何でも使える
 - 二値の符号にすれば殆どどんな媒体・技術でも使える
- インターネットの価値
 - 汎用デジタル通信網であることの意味
 - 突然それが現れたという衝撃

文字のデジタル表現

- 数値化された文字、とは？
 - あり得る文字にすべて番号を振る
 - 文字に番号を振って、文字列を番号列として表現
 - 番号付け = コード化(符号化)
- (例)
 - ABC = 1,2,3 とすれば 26 で足りる
 - abc = 27,28,29.. で 52 まで
 - 0,1,2 = 53,54 で 62 まで
 - 漢字はたいへんだが 6 万もあれば？

文字コード

- 文字番号表(この字を何番とするか)はいくつかある
 - 統一されていない(複数のフォーマットがある)
 - 言語の異なる相手とメールを交換すると？
- ASCII コード
- JIS漢字表
 - 第一水準、第二水準
 - JIS/EUC/Shift-JIS漢字コード

文字データの例

AB123 → “A” “B” “1” “2” “3”
65 66 49 50 51 (ASCII)

漢字 → “漢” “字”
180 194 187 250 (EUC)

- 標準枠の存在
 - 無限に大きな数字を書ける記録枠を用意したくない
 - 小さな桁数の枠をたくさん用意して、桁が足りない場合は並べて使う(工夫が重要)

標準枠としての Byte

- Byte
 - 慣例的に決まった 0-255までの256種類の値を入れられる枠 (8bit)
 - 255を超える値は二桁 (2Bytes) 使う
 - Byte is not 'Bite' (ガリっと噛む)
- ASCIIは 1 バイト
- 漢字は (普通は) 2 バイト
 - 「フロッピー1枚は新聞何枚に相当し、」

音楽CDは何バイトあるか？

- さまざまなもののバイト数
 - 広辞苑 (第二版)
 - 24字 x 50行 x 4段 x 2400ページ=11,520,000 字
 - 一文字 2 Bytesとして 23 Mega Bytes
 - 音楽CD
 - 44KHz x 65536段階(2Bytes) x 2ch = 176KB/sec
 - 176KB x 3600sec = 633,600 KB = 634MB

電子回路でのデータ表現

- コンピュータはスイッチの on/off で動作している
 - 数学的表現「二進法で動作している」
 - スイッチ一つ分、電線一本分、二進数一桁分のデータの枠を bit (binary digit) と呼ぶ
 - データ表現の最小単位のことを指す場合も
- 理由
 - 二値動作がもっとも機構的に単純 = 誤動作しにくい
 - そのぶんだのようなメディア・機構でも利用できた
 - 音はともかく、いま三値回路、四値回路があっても不思議ではない

電子回路でのデータ表現

- 1 Byte = 8 bits
 - 8bits が標準枠となった理由は？ (調べてみよ)
 - 4/8/12/16/32/36 と基準単位を 8 以外にとったものは幾らもある (工学系の人は PDP-11 を調べよ)
 - 1 Byte = 8 bits が今は普通
- 二進数表記
 - 中学生ごろの算数クラスを思い出して
 - 1 Bytes は 8 桁の二進数で表記される 11001011 という具合

データは bit へ

- デジタルデータは数値列として表現
- データは標準枠 Byte 単位に格納される
- 1 Byte = 8 bits
 - つまり 8 組の電子回路を単位に格納される
- 1 bit は on/off の二値
 - データは二値化されて電子回路に格納される
 - すなわち内部は二進法で数値処理が行われる
- デジタルデータが電気の on/off によって処理される姿が想像できただろうか？

数学的テクニック

- 処理対象 (データ) = 数値
 - 数学的なテクニックが使える
 - デジタル処理の重要な特性
 - アナログ処理でも駆使しているが、デジタル処理の方がはるかに柔軟性が高い
- 応用例：
 - 通信誤り訂正
 - 圧縮
- 単なる数列として処理する以上、数の計算を利用した加工がいくらかでも可能に

通信誤り訂正

- 間違いなく送る (記録する) 方法のひとつ
 - 送った元と先が同じであることを確認する方法
 - ノイズ対策
 - 多様な実現法: Checksum, CRC, 二度送るなど
- データのためのデータ
 - デジタル処理の柔軟さの根元
- より優れたデータ化が望まれる
 - そこに工夫の余地がある
- 様々な目的に応じた様々なフォーマットが

圧縮

- 表現次第でデータを短くすることはできる
 - 音楽CDの無音部分や絵の真っ白の部分を記録する効率的な方法はないか?
 - ほとんど聞こえない音を除いて記録してはどうか?
 - 同じ内容を記録するにしても、方法(フォーマット)は幾通りもある
- データを変換するという事
 - つまり数値を計算して別の数値を導き出すということ
 - デジタル処理の可能性の本質 (工夫の可能性)
 - 可逆と不可逆 (DVDは不可逆の圧縮)
 - 放送局は 300Mbps 程度で無圧縮、DV は 30Mbps 程度、DVD は 12Mbps 程度

まとめ

- データとフォーマットの関係
- 文字コード
- 標準枠としての Byte
- 二値電子回路のための bit
 - データは bit になって処理される
- データは数値 = 数学的加工処理
 - フォーマット変換、検証、圧縮、etc..
 - データのためのデータの存在

デジタルテレビ放送

- BSデジタル放送 2000 年 12 月スタート
 - 2003年2月末で世帯普及率 8.1% (社団法人 BSデジタル放送推進協会)
 - http://www.bpa.or.jp/news/2003_03/news8.1percent_030312.html
- 地上波デジタル放送 2003年末放送開始
 - 2003年2月9日 アナログ周波数変更スタート。青梅沢井局 (東京) で開始
 - <http://www.tokyo-dtv.org/> 地上デジタル放送
- 衛星放送のデジタル化が先行
 - 地上波は中継設備ごとデジタル化の必要がある
 - 衛星放送は局、衛星、受信者だけが対応すればよい

地上デジタルテレビ放送

- 総務省計画
 - 2001年7月、放送普及基本計画を変更
 - 地上デジタルテレビ放送開始時期は、
 - 関東・中京・近畿の三大広域圏は 2003年まで
 - その他の地域は 2006年まで
 - 高精細度テレビジョン放送 (HDTV)中心
 - 地上アナログテレビ放送は 2011 年までに終了
- まさに国策

地上デジタル放送への移行計画

- テーマ
 - 既存のアナログテレビ放送をすべて地上波デジタル放送へ変更
 - 2011年にはアナログ放送を終了する (一過性投資)
- 目的
 - 多チャンネル (従来より電波の使用帯域が少なくて済む)
 - HDTV への移行 (アナログでは失敗した)
 - デジタルデータ転送 (インターネット) との親和性
- 障害
 - 経済問題: 一億台の受信機 (TV, VTR) の置換を意味する
 - 技術問題: 新規に割り当てられる電波の周波数帯がない
 - 局側設備の更新など他にも問題はありますが、

周波数、周波数、周波数、、、

- デジタル放送を受信するために
 - 新規にデジタル受信機を購入する
- アナログ放送チャンネルの移動がたいへん
 - ユーザの受信機のチャンネル設定、アンテナを個別に変更
 - 変換器を介してデジタル放送を直接受信する場合も
- 個人にかかる経費は国費負担
 - 1800 億円を予定
 - 放送事業者、企業受信者などは自己負担

デジタル放送の利点

- 多チャンネル
 - 従来の 1 チャンネル幅に 3 チャンネル
 - 周波数帯の高効率利用
- HDTV
 - ノイズなどのない高画質
- これらは本当に「デジタル的」なのか？
 - アナログ技術でもできることでは？
 - 「量的向上」ではない「質的变化」は何か？
 - それを「デジタル的」と考えたい（そこにデジタル化の価値がある）

デジタル放送の利点

- マルチ放送
 - 1ch x 3 と 3ch 放送を混在させられる
 - 異なるフォーマットで送られてくるが、フォーマットを説明するデータを見て識別
 - データのためのデータが活かされている

NHK
<http://www.nhk.or.jp/digital/ground/service/index.html#q9>

デジタル放送の利点

- デジタル処理のために実現できたが、あまりデジタル的でないもの
 - 双方向？
 - データ放送？
 - 文字放送？
- もっとデジタル的な放送形態があり得たのではないか？

よりデジタル的に

- もっとデジタル的な形態があり得たのではないか？
 - フォーマットを解放したら
 - 縦横の画面比率を一定にする理由は（PC は既に自由ではないか）
 - もっと画質の低い番組で 10 分割したら
 - 超高画質だが 1 sec に一枚しか動かないものは
 - もっとデータのためのデータを活かさないか
- いずれ経路は電波という制約から離れる
 - 今の予定では 120Mbps 程度あればよいはず
 - じきに有線で可能になる
 - その時何が起きるか？

よりデジタル的に

- 受聴のスタイルが根本に変わる可能性がある
 - CoCoon : データのためのデータを活かしている
 - 過去にビデオが何をもたらしたか？
- 最後に放送局に残るものは何か
 - 120Mbps のデータを電波で配付する権利だけ
 - CoCoon 的ユーザからは単なる素材提供者に見える
- インターネット世界の人たちの視点
 - 価値のあるデータを、欲しいユーザにどう転送し、集金するかということをストレートに見ている
 - これこそデジタル化の価値

インターネット

- インターネットの価値
 - 汎用デジタル通信網であること
 - End to End で結ばれているということ

「放送局も、視聴者も、デジタル通信網の両端でデータ交換をする二者に過ぎない」という視点が現実