

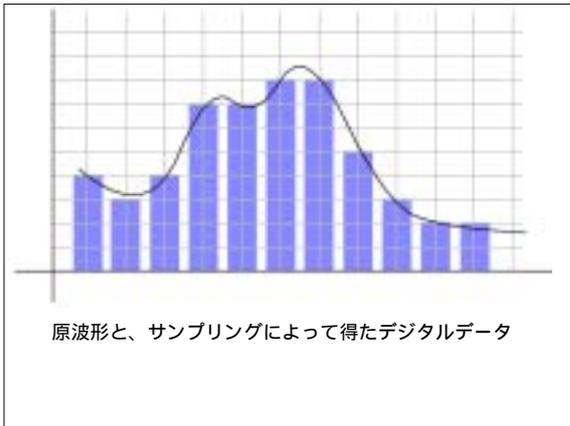
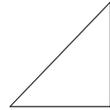
情報処理の概念

#3 デジタル表現 (続き) / 2002 (春)

一般教育研究センター 安田豊

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



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
- 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) :
 - 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

デジタル化による特長

- デジタル化 (数値表現化) の時点で発生するオリジナルとの相違 (ノイズとみなす)
- 伝達や記録、複製 (どれも実は同じこと) 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
- 作業の一部のデジタル化でもこの効果が得られる
 - M-Stageによる音楽の販売
 - ハリウッドのデジタル映画販売

デジタル化による特長

- 汎用の資源を利用できる
 - 数値を伝えればよいと言う点でデータの表現が汎用である
 - PicWalkはPHSとメモリーカード
 - さまざまな用途で使われる記録デバイス (CD, DVD etc..)
 - 通信経路はインターネットでまかなう
 - デジタル映画配信はDSL技術を使って
- インターネットの衝撃
 - すべての通信 (データ転送、データ流通) の経路をインターネットでまかなえる
 - これがインターネットという汎用デジタルネットワークが End to End で結ばれていることの本質
 - すべてのデジタルデータ交換が可能になった

データ

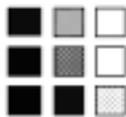
- ・ その実体は数値（記号）の列
 - ・ 音声：111,121,122,89,80,82,75....
 - ・ 静止画：10,240,22,30,34,80...
 - ・ 音声付き動画：12,33,45,1123,488...
- ・ これだけでは無意味
 - ・ 符号化ルールとデータは常に一体
 - ・ それがどんなものか
 - ・ どのようにして数値化したのか

符号化ルールと復元

- ・ データを元の「何か」に戻せるように
 - ・ 符号化にはルールが必要（でたらめでは駄目）
 - ・ 対応する復元ルール（逆関数）も必要
- ・ （ちなみに）復元だけ考えると
 - ・ 必ずしも符号化ルールの詳細を知る必要はない
 - ・ どのルールを使えば良いかが判ればよい
- ・ このルールがすなわちフォーマットを生む

フォーマット

- ・ 数え方を違えると全く違う数字列に
 - ・ 1-3-5,1-2-5,1-1-4
 - ・ 1-1-1,3-2-1,5-5,4
- ・ 数値化ルールと一致する複合化をしないと違う結果になる
 - ・ JPG, BMP, GIF etc...



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
 - ・ 計測、記述したルールを知らないと元に戻せない
- ・ 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
 - ・ 異なるアプリケーションでデータが扱えない理由
 - ・ 「互換性」の概念

スライド終了

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

- 文字番号表はいくつかある
 - 統一されていない
 - 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
- ISO2022-JPのコード表切り替え
 - コードを切り替える、というルール
 - さまざま複雑なフォーマットがあり得る

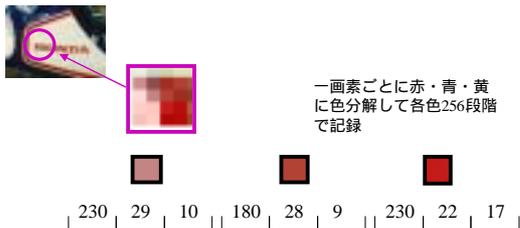
文字データの例

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）

画像データの例



Byte (通常先頭は大文字 B)

- 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
- 255 を超える値は二桁 (2 Bytes) 使う
- 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、」
 - CDについて計算してみる (次頁)

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 16bits x 2ch = 176KB/sec
 - 176KB x 3600sec = 633,600 KB = 634MB

bit/Byte/2進数

- ちょっと高校(中学?)の数学を思い出して
- 二進数
 - ドラム式時計を想像せよ
 - (バイクのメーター?)
 - 十進とはドラムに十種類(0-9)の記号がある
 - 二進とはドラムに二つしか数字が打ってない
- 十二支 = 12進数にも出来る
 - 今年はウシウシウマトラか？

bit (通常先頭は小文字 b)

- ・ 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 と呼ぶ
 - ・ 8bit が一つの区切りとなった
 - ・ 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 (工学系の人には PDP-11 を調べよ)
 - ・ 1 Byte = 8bits が今は普通

圧縮 (ちょっと話が戻って)

- ・ 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?
 - ・ 同じ内容を記録する方法は幾通りもある
 - ・ エラー訂正 (後述)
 - ・ データを変換するという事
 - ・ 内容を残したまま違うフォーマットにする

エラー訂正

- ・ 同じ内容を送る方法のひとつ
 - ・ ノイズ対策強化
 - ・ 量子化レベルを超えたノイズでも復元可能
 - ・ Checksum, CRC, 二度送ったって良い
- ・ 単なる数列として処理する以上、数学を利用して可能な加工はいくらでも可能
- ・ このような情報の付加を「冗長」という
- ・ 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

工夫

- ・ エンジニアリングの本質
- ・ コンピュータは工夫の集積箱
 - ・ 必ず理解できるはずのもの
 - ・ 僅か半世紀ほどの歴史
 - ・ 物理学や数学などより具体的に学びやすい (はず)
 - ・ そのかわり変化が激しい
 - ・ 広い視野をもつことで対抗

西陣織の紋図と紋紙

- ・ サンプルによるデジタル化そのもの
 - ・ 紋紙の穴ひとつひとつが 1 bit に相当する
- ・ 自動処理機械として極めてコンピュータ的
 - ・ 1ステップの動作仕様が書かれている
 - ・ 単純な機能を組み合わせ、複雑な結果を得る
 - ・ コンピュータも同じ
 - ・ CPU の処理能力をよく Hz で表現する
 - ・ なぜ僕の PC は一気に 100GHz にならないのだろう?
 - ・ ステップを踏んで処理をするということ

歴史

- ・ ジャカード(F)の自動織機は1805頃
 - ・ 1781 にワットの蒸気機関 (産業革命)
 - ・ もう一つの革命の入り口がそこに
- ・ ホレリス
 - ・ 1890ホレリス(US)がUS国勢調査の統計処理をパンチカードを利用した加算機を用い、10年かかる仕事を2年半で仕上げた。
 - ・ 1896 設立したホレリスの会社はIBMへと
 - ・ 1935には独禁法で司法省と戦うほど成長
- ・ その後のコンピュータに直結する道すじの一つ

歴史

- ・ 時間をかけた工夫の集積体としてのコンピュータ
 - ・ 数学的（もしくは理論的）裏付けと同時に理解する
- ・ 変化のただなかで
 - ・ 自己の視点の正しさを自分でチェック可能
 - ・ 「昔、映画はフィルムだった」と言える可能性大