

1

Presentation Formatting

COS 460/540

2

End to End Data

- Data Formats
- eXtensible Markup Language
- Multimedia Data

3

Data Formats

- Basic data types
- Complex types and data
 - “records”, audio, video, ...
- Sequences

Complex Types

- Compression
 - ...to reduce bandwidth needs
- Error Correction
 - ...to increase reliability

Transmitting Data

- Encoding
 - ...from model to network
- Decoding
 - ...from network to model

XML

eXtensible Markup Language

- Data and Tags/Markup (XML)
- Schema description of documents (XSD)

7

```

<?xml version="1.0"?>
<catalog>
  <book id="bk101">
    <author>Gambardella, Matthew</author>
    <title>XML Developer's Guide</title>
    <genre>Computer</genre>
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at applications
    with XML.</description>
  </book>
  <book id="bk102">
    <author>Ralls, Kim</auth

```

8

XML

- Based on Web Technologies
- Data and Markup are TEXT
- XML is a “framework”
- Nested tags/values
- Sequences of tags/values

9

```

<xsd:schema xmlns:xsd="http://www.w3....XMLSchema"
  targetNamespace="urn:books"
  xmlns:bks="urn:books">
  <xsd:element name="books" type="bks:BooksForm"/>
  <xsd:complexType name="BooksForm">
    <xsd:sequence>
      <xsd:element name="book"
        type="bks:BookForm"
        minOccurs="0"
        maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="BookForm">
    <xsd:sequence>

```

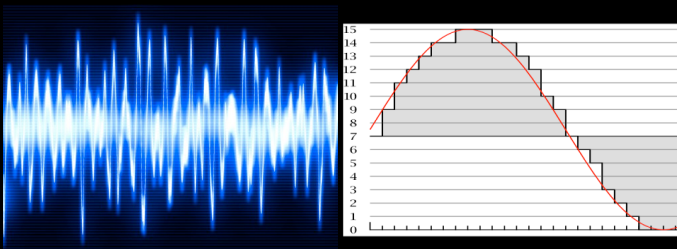
XSD

- Defines valid XML Documents
- Written in XML
- Basic types: integer, string, boolean
- Complex types: nesting, sequences
- Namespaces to avoid name conflicts

Multimedia Data

- The nature of multimedia data
- Compression
 - Lossless (for data)
 - Lossy (for images, video, audio)

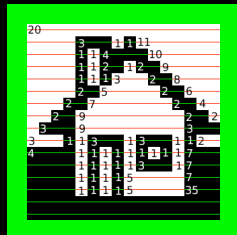
Audio



Sampling

- Sampling (time)
- Quantization (quantity, e.g. amplitude)

Images & Video



video is a sequence of images

That's a lot of data!

$$1080 \times 1920 \times 24 = 50\text{Mb}$$

$$24\text{fps} = 1.2\text{Gbps}$$

Lossless Compression

All the data are
important

- Run Length Encoding
- Differential Pulse Code Modulation

Run Length Encoding

AAABBCDDDDDDAAAAABCCC 21B



12B

11B

DPCM

AAABBCDDDDDDAAAAABCCC 21B

(differential)  12B

(delta)  11B
(RLE delta) 10B

Huffman Code

Morse Code

A ··	J ····	S ···	1 ·····
B ····	K ···	T =	2 ·····
C ····	L ····	U ··	3 ·····
D =·	M =	V ····	4 ·····
E ·	N =	W =·	5 ·····
F ····	O =·	X ····	6 ·····
G ···	P ····	Y ····	7 ·····
H ····	Q ····	Z ····	8 ·····
I ··	R ···		9 ·····
			0 ·····

LZW - Dictionary

industrial workers of the world, an international industrial labor union that was organized in C in 1905 and disintegrated after 1920. *Abbr.*: I.W.W., I.

in-dus-tri-ous (in dus'trē əs), *adj.* 1. hard-working; diligent. 2. *Obs.* skillful. [*< L industrius, OL indostrius, disputed origin*] —**in-dus'tri-ous-ly**, *adv.* —**in-dus'tri-ous-ness**, *n.* —*Syn.* 1. assiduous, sedulous, energetic, busy. —*Ant.* 1. lazy, indolent.

in-dus-try (in'da strē), *n., pl. -tries* for 1, 2. 1. the system of manufacturing or technically productive enterprise in a particular field, often named after its principal product. 2. any general business field. 3. trade or manufacture in general. 4. owners and managers collectively. 5. system of work or labor. 6. assiduous activity at any work or occupation. [*ME industrie < L industrius, from ind-, dev- + dusterius INDUSTRIOUS*] —*Syn.* 6. effort, endeavor, devotion.

Indus val/ley civiliza/tion, an ancient civilization that flourished in the Indus River valley in India, about 1500 B.C. Also called **Indus civilization**.

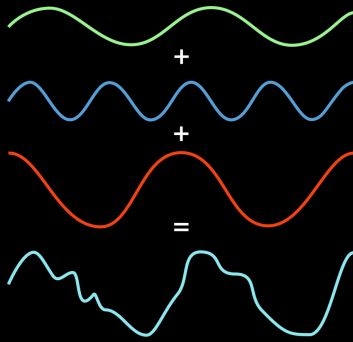
in-dwell (in dwell'), *v., -dwelt, -dwelling.* 1. to inhabit. 2. to possess (a person), as a principle, motive force, etc. —*i.* 3. to dwell. 4. to abide within, as a person, etc. [*ME indwelle(n) < ind- + dwell*]

Lossy Compression

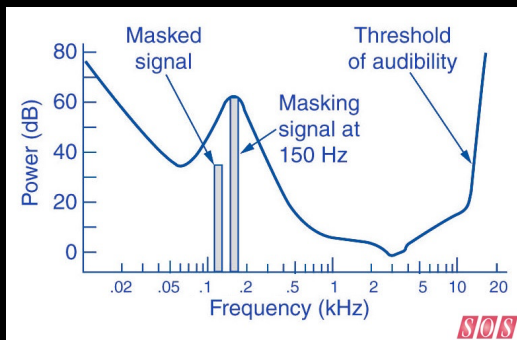
All the data are
NOT important

- Single image compression
- Stream compression

MP3 (audio)



Signal Reduction (masking)



CBR vs VBR (constant vs variable bitrate)

24Kbps = Spoken word (telephone)
 48–64Kbps = AM Radio
 128Kbps = reasonable for car-radio, falls off over 16KHz (cymbals)
 192KBps = 'near CD quality'
 $\geq 256\text{Kbps}$ = identical to original up to about 18KHz

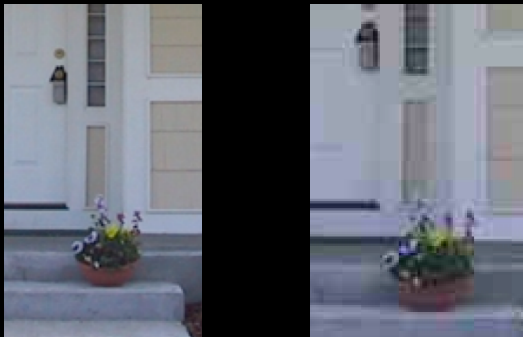
Bitrate?

- 24Kbps = Spoken word (telephone)
- 48–64Kbps = AM Radio
- 128Kbps = reasonable for car-radio, falls off over 16KHz (cymbals)
- 192KBps = 'near CD quality'
- $\geq 256\text{Kbps}$ = identical to original up to about 18KHz

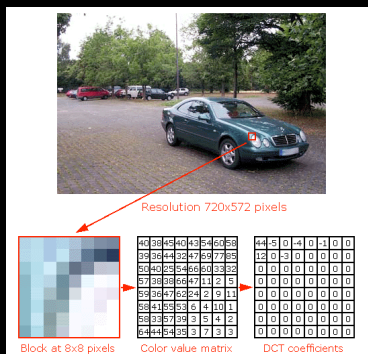
CBR vs VBR

- Constant
 - Same bitrate throughout the stream
- Variable
 - bitrate changes based on content analysis

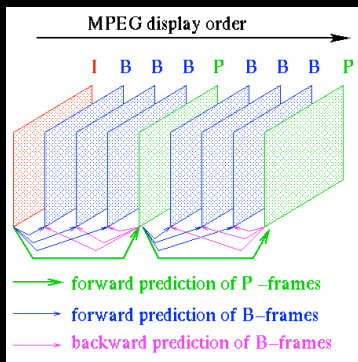
Images & Audio



JPEG



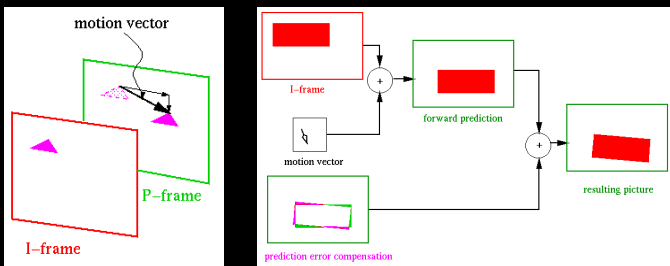
MPEG



28

https://vsr.informatik.tu-chemnitz.de/~jan/MPEG/HTML/mpeg_tech.html

Prediction



29

End

Presentation Formatting
XML

30