

---

# 第一章 绪论

**多媒体的定义：**以数字化为基础，能够对多种媒体信息进行采集、编码、存储、传输、处理和表现，综合处理多种媒体信息并使之建立起有机的逻辑联系，集成为一个系统并能具有良好交互性的技术。

**5种媒体：**感觉媒体，表示媒体，显示媒体，存储媒体，传输媒体

**多媒体的关键特性：**

**多维化：**计算机处理媒体信息的多样化，使人与计算机之间的交互不再局限于顺序的、单调的、狭小的范围，而有充分自由的余地。

**交互性：**人、机对话，是多媒体技术的关键特征。在多媒体系统中，除了操作上控制自如之外，在媒体的综合处理上也可以随心所欲。

**集成性：**将各种不同的媒体信息有机地进行同步组合，形成一个完整的多媒体信息；把不同的媒体设备集成在一起，形成多媒体系统。

多媒体是技术与应用发展的必然产物

多媒体改善了人类的交流：

**使得人类的信息处理手段得到加强：** 计算能力；存储；高速通信网

**计算机的工作方式：** 无法自由收集信息和表达信息；交互过程中信息的转换

**人机交互形式：** 计算机—计算机；人—人（通过计算机）；人—计算机；计算机—人

**多媒体：** 使计算机理解信息的含义；用户接口工作；人的思维—>机器指令

**多媒体缩短了人类传递信息的路径：** 1、强大数据组织和构造能力 2、高效的算法和高速的网络通信 3、“量体裁衣”

**多媒体技术的研究：** 一、多媒体技术： 基本技术层面的内容

二、多媒体系统： 多媒体系统的构成与实现

**多媒体技术：** 包括

多媒体技术基础：媒体研究、数据压缩

多媒体软硬件平台技术

多媒体操作系统技术

多媒体信息管理与处理技术

多媒体通信与分布应用技术

**MATLAB (Matrix Laboratory):**

用一个简单命令求解线性系统

$$3x_1 + x_2 - x_3 = 3.6$$

$$x_1 + 2x_2 + 4x_3 = 2.1$$

$$-x_1 + 4x_2 + 5x_3 = -1.4 \quad \text{Code:} \quad A = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 2 & 4 \\ -1 & 4 & 5 \end{bmatrix}; \quad b = \begin{bmatrix} 3.6 \\ 2.1 \\ -1.4 \end{bmatrix}; \quad x = A \backslash b$$

$$x = 1.4818 \quad -0.4606 \quad 0.3848$$

用简短命令计算并绘制在 $0 \leq x \leq 6$ 范围内的

的 $\sin(2x)$ 、 $\sin x^2$ 、 $\sin^2 x$

$x = \text{linspace}(0, 6)$

$y_1 = \sin(2*x), y_2 = \sin(x.^2), y_3 = (\sin(x)).^2;$

$\text{plot}(x, y_1, x, y_2, x, y_3)$

## 第二章 媒体与媒体技术

### 2.1 媒体的种类和特点

#### 常见的媒体元素：

**文本：**非格式化文本：只有文字信息，没有其他任何有关格式的信息

格式化文本：带有各种文本排版信息等格式信息的文本文件

ASCII: American Standard Code for Information Interchange

**图形(Graphic)：** 特征：图形是对图象进行抽象的结果（人工或自动）

图形的矢量化使得有可能对图中的各个部分分别进行控制（放大、缩小、旋转、变形、扭曲、移位等）

图形的产生需要计算时间

**图像(Image)：**用数字点阵方式表示的场景画面

分辨率：屏幕分辨率，图像分辨率，像素分辨率

图像灰度：每个图像的最大颜色数，屏幕上每个像素都用一位或多位描述其颜色信息。 24 位为真彩色

**图像文件大小：**用Byte(字节)为点阵表示图像文件大小时描述方法为： $(高*宽*灰度位数)/8$ 。

**视频 (Video)：**由连续的画面组成。这些画面以一定的速率连续地投射在屏幕上，使观察者具有图像连续运动的感觉。

#### 视频的制式

- PAL制（625/50）每秒25帧，水平扫描线为625条，水平分辨率240~400个像素，隔行扫描。62年诞生于德国，应用于中国、新加坡、欧洲地区等（Pal-B、D、G、H、I、N、NC）。

- NTSC制（525/60）每秒30帧，水平扫描线为525条，水平分辨率240~400个像素，隔行扫描。53年诞生于美国，应用于美国、日本、台湾等（NTSC-M、NC、Japan等）。

- SECAM制（625/50）每秒25帧，水平扫描线为525条、水平分辨率625。由法国人提出，应用于俄罗斯、法国、非洲地区等。

**视频的数字化：**指在一段时间内以一定的速度对视频信号进行捕获并加以采样后形成数字化数据的处理过程。

视频主要技术参数见书P12

**动画 (Animation)：**动画是运动的图形，其实只是一幅幅静态图形的连续播放。

造型动画 和 帧动画（P12）

**音频 (Audio)** 波形声音、语音、音乐 主要因素：采样频率、采样精度、通道数

**媒体的种类：**视觉类媒体、听觉类媒体、触觉类媒体、嗅觉和味觉媒体

在多媒体技术中 主要是前三种媒体，即表示媒体

#### 媒体的性质和特点：

- 各种媒体具有不同特点和性质 1、媒体是有格式的 2、不同媒体表达信息的特点和程度各不相同 3、媒体之间可以相互转换 4、媒体之间的关系也具有丰富的信息

- 媒体具有空间性质 1、表现空间 2、媒体按相互的空间关系进行组织

3、视觉空间、听觉空间和触觉空间这3者既相互独立又需要相互结合

- 媒体的时间性质 1、表现需要时间 2、媒体在时间坐标轴上的相互关系

- 媒体的语义 1、媒体的语义是有层次的 2、抽象的程度不同，语义的重点也就不同

- 媒体结合的影响 1、媒体结合是多层次的 2、媒体结合有利于信息接受和理解

- 隐喻 - 交互的概念模型，也称心智模型

### 2.2 听觉媒体技术

#### 2.2.1 声音心理学

听觉特性：

- 
- 1、**等响曲线** 描述响度与频率和强度的关系 **横坐标为：频率 纵坐标为：声压级(d)**
  - 2、**掩蔽**：一种声音的出现可能使另一种声音难以听清，声音的掩蔽效应可用于声音的压缩
  - 3、**临界频带**：在频率的某一临界区里，各种声音的强度是相互作用的，合成声音的响度由这些频率共同决定
  - 4、**相位**：相位的确定对于多声道声音系统的设计非常重要 应用：回升的消除、会议系统的声音设计。
  - 5、**自然声音的时变现象**：声音的音调分成三个区域：起始区、稳定状态区、延迟区
  - 6、**听觉空间**：人耳可听到来自各个方向的声音，并用不同的因素来判定声源的位置。声源的位置不论对于增进人们的感受还是增进对声音的理解，都是非常重要的，通过声音的精确再现，就可以造出听觉空间
  - 7、**听觉的频谱特性**：声音是时间函数，通过傅里叶变换可做出其频谱图，人耳对频谱成分的波峰和波谷是非常敏感的，基频改变，人耳是很敏感的。例如：快进的录像，音调会发生变化。
  - 8、**声音的心理模拟**：通过人工真实的方法，可以对视觉空间的景物进行再造或虚构，同样也可以对听觉空间的声音进行心理的模拟，这就是所谓的可听化（**audiolization**）。用声音可以表达出一些声音的效果。

## 2.2.2 音频的数字化和符号化：

### 1、音频的数字化和再现

模拟信号→采样→量化→编码→数字信号

PCM:Pulse Code Modulation 脉冲编码调制 采样频率越高，声音质量就越接近原始声音，所需的存储量也就越大。

声音的数据量=（采样频率\*每个采样位数\*声道数\*录音时间s）/8 （Byte/s）

一首5分钟CD音乐光盘引致的歌曲，文件大小是多少？ Key:50.47MB（44.1KHz, 16位, 立体声）

### 二、声音的符号化

音乐符号化：MIDI (Music Instrument Digital Interface) 乐器数字接口

MIDI消息，是指乐谱的数字描述

任何电子乐器，只要有处理MIDI消息的微处理器和合适的硬件接口，就

构成了一个MIDI设备。当一组MIDI消息通过音乐合成芯片处理时，合成器能解释这些符号并且产生音乐。

MIDI 实质上是由 MIDI 控制器（或 MIDI 文件）产生的指示电子音乐合成器要做什么、怎么做（如演奏某个音符、加大音量、生成音响效果）的一套标准指令。MIDI 不是声音信号，在 MIDI 电缆上传送的不是声音，而是动作指令。

单个物理MIDI通道分为16个逻辑通道，每个逻辑通道可指定一种乐器。MIDI键盘可设置在这16个通道中的任何一个，MIDI合成器可以被设置在指定的通道上接受。

语音符号化：对语音的符号化实际上就是对语音的识别，将语音转变为字符，反之也可以将文字合成语音。

## 2.2.3 音频媒体的三维化处理

### 1. 三维虚拟声空间

Three Dimensional Virtual Acoustic, 3DVA是指用一定的声音设备人为地产生出来的具有空间位置信息的声音空间。三维听觉的使用明显地依赖于用户对听觉空间中各种信息源的定位能力。

### 2. 3DVA的基本理论

人类感知声源位置的最基本的理论是双工理论(Duplex Theory)：

两耳间声音的到达时间差（因距离原因所致）ITD (Interaural Time Differences)；

两耳间声音的强度差（因信号衰减所致）IID (Interaural Intensity Differences)。

- 大脑根据IID和ITD分别对高频与低频声音定位

## 2.3 视觉媒体技术 (P24)

### 1. 视觉的心理特征

同声音心理学相似，与视觉相对应的光学物理性质和心理知觉也是截然不同的。虽然光的物理特性与心理知觉有关，但它们的关系并不是线性的。对光的色调和亮度的感觉不仅和它的频率和强度有关，而且还和它出现的背景有关，和同时出现的周围光有关。

## 2. 视觉特性

A. **亮度** 亮度是人眼对光强度的感受。一个物体的亮度不仅跟目标的**物理强度**有关，而且与**周围的背景**有关。对图像的处理最重要的是**亮度的差别**。

1. **马赫带** 在明亮到黑暗的过渡部分，有一条特别暗和一条特别亮的光带，它反映了感觉上的物理量和心理量的不同

2. **等亮曲线** 反映了视觉在亮度上与波长的关系

3. **掩蔽现象**

## B. 视觉的时间特性

人眼在观察景物时，光信号传入大脑神经，需经过一段短暂的时间，光的作用结束后，视觉形象并不立即消失，这种残留的视觉称“**后像**”，视觉的这一现象则被称为“**视觉暂留**”。视觉暂留：**0.1-0.4 秒** 具体应用：电影的拍摄和放映

## C. 彩色

**三要素**：（任何一种颜色都可以用这三要素来确定）

1. **色调**：颜色的种类，红、橙、黄、绿、青、蓝、紫等

2. **亮度**：颜色的相对明暗程度，0%（黑色）-100%（白色）

3. **饱和度**：表示颜色的深浅，灰色的成分所占的比例，0%（灰色）100%（完全饱和）

**RGB 颜色模型** • 理论上绝大部分可见光谱都可用红、绿和蓝（**RGB**）三色光按不同比例和强度的混合来表示。  
**颜色C=R(红色的百分比)+G(绿色的百分比)+B(蓝色的百分比)**

• RGB模型称为**相加混色模型**，用于光照、视频和显示器。例如，显示器通过红、绿和蓝荧光粉发射光线产生彩色。

**CMYK 颜色模型** • 在理论上，绝大多数颜色都可以用三种基本颜料（青色cyan、品红magenta、和黄色yellow）按一定比例混合得到。

• 理论上，青色、品红和黄色三种基本色素等量混合能得到黑色。但实际上，因为所有打印油墨都会包含一些杂质，这三种油墨混合实际上产生一种土灰色，必须与黑色（**K**）油墨混合才能产生真正的黑色，所以再加入黑色作为基本色形成**CMYK**颜色模型。

• CMYK模型称为**相减混色模型**。

屏幕上显示的图像，都是RGB模式

印刷品上的图像，都是CMYK模式

**HSL 颜色模型** • 在HSL模型中H: Hue, 色调 S: Saturation, 饱和度 L: Lum, 亮度  
模型主要是面向设备的，而HSL模型更容易被人理解和控制。

• RGB模型和CMYK

表示颜色需要考虑三种**心理属性**：**色调、亮度和饱和度**。

**色调**就是通常意义下的彩色，它随波长的变化而变化，反映颜色的基本特性。

**亮度**是光作用于人眼所引起的明亮程度的感觉，与光强有关。

**饱和度**是代表颜色的深浅程度。色调和饱和度统称为**色度**。

## D. 注视点和视野范围

**注视点**：黑白交接时隐时现不规则

**视野范围**：左右视角约为180度 上下视角约为60度 视力好的部位2到3度

### 2.3.2 模拟视频原理

**模拟电视**：以模拟电信号记录，依靠模拟调幅传播，以盒式磁带录像机存储的

**数字电视（Digital TV）**：从演播室到发射、传输、接收的所有环节都是使用数字电视信号或对该系统所有的信号传播都是通过由0、1数字串所构成的数字流来传播。

**分辨率**表现的是电视系统中重现场景细节的能力。水平扫描线所能分辨出的点数称为**水平分辨率**。一帧中垂直扫描的行数称为**垂直分辨率**。垂直分辨率和每帧中的扫描线有关，扫描线越多，分辨率就越高。

**2.3.4 视觉媒体的三维立体显示**：3D电影利用人眼**视差** 模拟双眼：1. 头盔显示器2. 分别拍摄两种图像，戴上眼镜观看

## 第三章 多媒体数据压缩

### 数字化声音：

一曲3分钟的CD音乐（44.1KHz，16bit）所占的存储空间是多少？

$$44100 \times 16 \times 2 \times 180 / (8 \times 1024 \times 1024) = 31\text{MB}$$

采用mp3格式压缩编码（压缩比15：1），则为2MB

### 数字化视频图像：

一张分辨率为1024 × 768，颜色深度为24的图像所占的存储空间是多少？

$$1024 \times 768 \times 24 / (8 \times 1024 \times 1024) = 2.25\text{MB}$$

用jpeg格式压缩编码（压缩比10：1）则为225KB

一小时1024×768的视频所占的存储空间呢？

$$2.25 \times 25 \times 3600 = 202500\text{MB}$$

采用MPEG2压缩编码（压缩比25：1）则为8100MB

**数据压缩核心问题：**利用最少时间和最小空间，传输和保存多媒体数据。

**数据压缩的必要性：**多媒体“数据爆炸”，如果不进行数据压缩，传输和存储都难以实用化

**数据压缩的可能性：**1、人类不敏感的因素 2、数据冗余（P38）

a. **空间冗余——图像的空间连贯性** 静态图像中存在的最主要的一种数据冗余

一幅图像表面上个采样点的颜色之间往往存在着空间连贯性，基于离散采样来表示物体表面颜色的像素存储方式可利用空间连贯性，达到减少数据量的目的。

b. **时间冗余——视频与动画画面间的相关性** 一组连续的画面之间往往存在着时间相关性

运动图像一般为位于一时间轴区间的一组连续画面，其中的相邻帧往往包含相同的背景和移动物体，只不过移动物体所在的空间位置略有不同，所以后一帧的数据与前一帧的数据有许多共同的地方，这种**共同性**是由于相邻帧记录了相邻时刻的同一场景画面，所以称为时间冗余。

c. **信息熵冗余——编码冗余，数据与携带的信息**

d. **视觉冗余——视觉敏感度和非线性感觉** 人类的视觉系统由于受生理特性的限制，对于图像场的主意是非均匀的，人对细微的颜色差异感觉不明显。

e. **听觉冗余——听觉敏感度** 人耳对不同频率的声音的敏感性是不同的

f. **知识冗余——凭借经验识别** 图像的理解与某些基础知识有关

g. **结构冗余——规则纹理、相互重叠的结构表面**

**数据压缩分类：**A. 按原始数据是否有损失：1. 有失真编码 2. 无失真编码

B. 按编码原理分类：1. 预测编码 2. 变换编码 3. 统计编码 4. 分析-合成编码 5. 混合编码

### 数据压缩技术的性能指标：

(1) **压缩比** 设n1和n2分别代表用来表达相同信息的2个数据集合中的信息载体单位的个数。 压缩率（压缩比）：——描述压缩算法性能  $CR = n1 / n2$

其中，n1是压缩前的数据量，n2是压缩后的数据量。

$$\text{相对数据冗余：} RD = 1 - 1/CR \quad \text{例：} CR=20; RD = 19/20$$

(2) **图像质量** 图像保真度 a. 无失真编码/有失真编码 b. 描述解码图像相对于原始图像的偏离程度 c. 对信息损失的测度

主观保真度准则 主观测量图象的质量，因人而异，应用不方便

客观保真度准则 用编码输入图与解码输出图的某个确定函数表示损失的信息量，便于计算或测量

(3) **压缩解压速度** 静态图像：要求不是很严格 动态视频：必须有较高的帧速（至少15帧/秒）

### ◆ 3.2 常用的数据压缩编码方法

**预测编码：**利用前面的一个或多个信号对下一个信号进行预测，然后对**实际值和预测值的差**进行编码。



原理：利用前面一个或多个像素值对当前待编码像素值进行预测，对差值而不是原像素值编码，在差值小于原始值时可以得到较低码率。

(1) DPCM (差分脉冲编码调制)

(2) ADPCM编码(自适应差分脉冲编码调制) (P43)

- PCM(Pulse Code Modulation)，原始的模拟信号经过时间采样，然后对每一样值进行量化，作为数字信号传输。

如某灰度图像PCM编码为：

129、128、127、127、126、126、125、124、123、122 预测方法： $f'(i, j) = f(i-1, j-1) - 1$

预测值：129、128、127、126、125、124、123、122、121、120……

差值信号：0、0、0、1、1、2、2、2、2、2 对差值信号进行量化、编码、发送，可以减小存储位数。

- DPCM不对每一样值都进行量化，而是预测下一样值，并量化实际值和预测值之间的差。
- DPCM是基本的编码方法之一，在大量的压缩算法中被采用，比如JPEG的DC分量就是采用DPCM编码的。

### 3.2.2 变换编码(P43)

将在空间域里描述的图像，经过某种变换（通常采用正交变换），在频域（变换域）里进行描述，降低图像的相关性，实现数据压缩。

**基本思想：**先对信号进行某种函数变换，将信号的表示方法从一种信号空间变换到另一种正交矢量空间，从而产生一批变换系数，然后再对这些系数进行编码

(1) **最佳变换（K-L变换）** 数据压缩主要是去除信号的相关性。若考虑到信号存在于无限区间上，而变换区域又是有限的，那么表征相关性的统计特性就是协方差矩阵。

**K-L变换的物理意义：**K-L变换实质上是作坐标系的转换，尽量让向量落在最少的坐标轴上或其周围，从而只用较少的变换系数就可以恢复出质量不错的图像，压缩效率比较高，均方误差小。缺点：图像变化后变换核矩阵也要相应变换，求解过程比较复杂，没有快捷方法。

(2) **离散余弦变换（DCT变换）** 如果变换后的协方差矩阵接近对角矩阵，该类变换称**准最佳变换**，典型的有DCT、DFT、WHT、HrT等。其中，最常用的变换是离散余弦变换DCT。

**离散余弦变换(Discrete cosine Transform)简称DCT。**任何连续的实对称函数的傅里叶变换中只含余弦项，因此余弦变换与傅里叶变换一样有明确的物理量意义。DCT是先将整体图像分成 $N \times N$ 像素块，然后对 $N \times N$ 像素块逐一进行DCT变换。由于大多数图像的高频分量较小，相应于图像高频成分的系数经常为零，加上人眼对高频成分的失真不太敏感，所以可用更粗的量化，因此传送变换系数所用的数码率要大大小于传送图像像素所用的数码率。到达接收端后再通过反离散余弦变换回到样值，虽然会有一定的失真，但人眼是可以接受的。

**为什么变换编码是一种有损编码？**变换本身是可逆的，因而其也是一种无损技术。然而，为了取得更满意的结果，某些重要系数的编码位数比其他的要多，某些系数干脆就被忽略了。这样，该过程就成为有损的了。

### 3. 统计编码 (1) 行程编码 (2) LZW编码 (3) 霍夫曼编码 (4) 算术编码

(1) **行程编码(Run Length Coding, RLC)** 将图像灰度值 $f(x, y)$ ，映射为整数对，用(A, B)表示。A表示灰度值，B表示具有该灰度值的连续像素的点数。E.g 88866666→(8, 3)、(6, 5)；

**行程编码特点：**a. 直观，经济；b. 是一种无损压缩；c. 压缩比取决于图像本身特点，相同颜色图像块越大，图像块数目越少，压缩比越高。d. 适用于计算机生成的图像，例如BMP、TIF等，不适于颜色丰富的自然图像。

**RLE 编码——Run Length Encoding**

概念：行程：具有相同灰度值的像素序列。

编码思想：去除像素冗余。

用行程的灰度和行程的长度代替行程本身。

举例说明：

例：设重复次数为 iC，重复像素值为 iP

aaaa bbb cc d eeeee ffffffff

编码为: iCiP iCiP iCiP (共22\*8=176 bits)  
 编码前: aaaaaaabbbbbcccccccc 4a 3b 2c 1d 5e 7f  
 编码后: 7a6b8c (共12\*8=96 bits)  
 压缩率为: 96/176=54.5%

(2) LZW编码 J. Ziv和A. Lempel在1978年首次发表了介绍第二类词典编码算法的文章。在他们的研究基础上，Terry A. Welch在1984年发表了改进这种编码算法的文章，因此把这种编码方法称为LZW(Lempel-ZivWalch)压缩编码。

LZW编码是围绕称为词典的转换表来完成的。

(3) 霍夫曼编码（作业）

原理：在哈夫曼编码中，对于出现概率大的信源符号编以短字长的码，对于出现概率小的信源符号编以长字长的码。如果码字长度严格按照符号概率大小的相反顺序排列，则平均的码字长度一定小于按任何其它字长顺序排列方式得到的码字长度。

编码步骤: A. 缩减信源符号数量 将信源符号按出现概率从大到小排列，然后结合

B. 对每个信源符号赋值 从（消减到）最小的信源开始，逐步回到初始信源

霍夫曼编码结果：

平均长度:  $L_{avg} = \sum_{j=1}^j l(a_j)P(a_j)$      信息熵:  $H(u) = - \sum_{j=1}^j P(a_j) \log_2 P(a_j)$

编码效率:  $\eta = \frac{H(u)}{L_{avg}}$      压缩比: 编码后位数/编码前位数

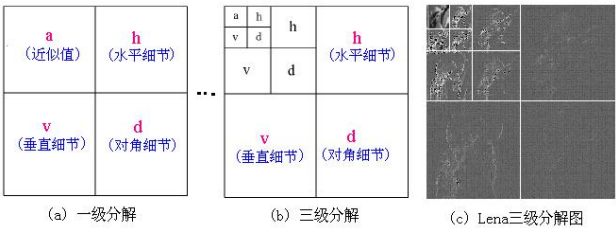
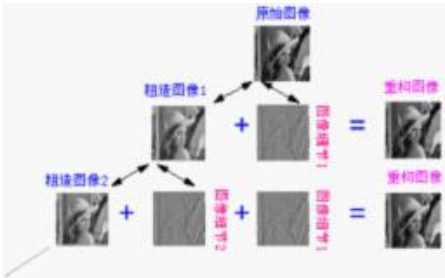
哈夫曼编码方法：

- (1) 哈夫曼方法构造出的码并不是唯一的
- (2) 哈夫曼编码对于不同的信息源概率分布，其编码效率不同

4. 分析—合成编码

原理：通过对原始数据的分析，将其分解成一系列更适合于表示的“基元”或从中提取出具有本质意义的参数，对这些基元或者参数进行编码。

分析-合成编码：(1) 量化编码 (2) 小波变换编码（P44）(3) 分形图像编码 (4) 子带编码



◆ 3.3 音频压缩标准

交互式多媒体音频系统： 输入：对自然界声音的数字化

输出：数字化声音还原成模拟音或系统生成的虚拟声音。

音频信息在编码技术中通常分成两类来处理，分别是语音和音乐，各自采用的技术有差异。

- 语音编码技术-波形编码：基于声音波形预测的编码技术。在时域上处理，利用声音波形相邻样值间相关性压缩冗余数据，力图使重建语音波形保持原始语音信号形状。

特点：将语音信号作为一般的波形信号来处理，具有适应能力强、语音质量好等优点，但压缩比偏低。

- 语音编码技术-参数编码：利用语音信息产生的数学模型，提取语音信号的特征参量，并按照模型参数重构

音频信号，只能收敛到模型约束的最好质量上。

**特点：**压缩比高，但重建音频信号质量较差，自然度低，适用于窄带信道的语音通讯，比如军事通讯、航空通讯等。

- **语音编码技术-混合编码：**结合波形和参数编码方法，基本原理是合成分析法：得到综合滤波器和最佳激励过程称为**分析**，用激励和综合滤波器合成语音过程称为**综合**

**特点：**把参数编码和波形编码的优点结合在了一起，使得用较低的数据率产生较好的音质成为可能。

- **语音编译器：**波形编译器(waveform codecs) 话音质量高，但数据率也很高  
音源编译器(source codecs) 数据率很低，产生的合成话音的音质有待提高  
混合编译器(hybrid codecs)使用音源编译码技术和波形编译码技术数据率和音质介于它们之间
- **音乐编码技术-自适应变换编码：**利用正交变换，把时域信号变换到另一个域，去掉相关，变换域系数能量集中在一个较小范围，对变换域系数最佳量化后可以实现码率压缩。

理论上的最佳量化一般很难达到，通常采用自适应比特分配和自适应量化技术来对频域数据进行量化。

- **音乐编码技术-熵编码：**按照信息出现概率的分布特性，在编码过程中不丢失信息量，即保存信息熵，是一种无损数据压缩编码。

根据信息论原理，可以找到最佳数据压缩编码的方法，数据压缩的理论极限是信息熵。常用的有**哈夫曼编码**和**算术编码**。

**音频压缩编码标准：**

(1) ITU的基准标准. (2) GSM标准. (3) CTIA标准. (4) NSA标准 (5) MPEG标准

## ◆ 3.4 图像和视频压缩标准

### 3.4.2 静止图像压缩标准 (P51)

JPEG 算法在1992 年被确定为JPEG 国际标准，是国际上彩色、灰度、静止图像的第一个国际标准，是ISO制定的。JPEG 标准是一个适用范围广泛的通用标准。它不仅仅适于静态图像的压缩，电视图像序列的帧内图像的压缩编码也常常采用JPEG 压缩标准。

**颜色转换：**

**RGB颜色空间：**每个像素以红、绿、蓝三个分量表示

**YUV颜色空间：**Y为亮度信号，U、V为色差信号 (B-Y、R-Y)

我国和德国电视系统采用的PAL-D制式

**YIQ颜色空间：**美国电视系统采用的NTSC制式

**颜色转换公式：**

$$Y = 0.299R + 0.587G + 0.114B \quad U = -0.169R - 0.332G + 0.5B \quad V = 0.5R + 0.419G - 0.081B$$

$$Y = 0.299R + 0.587G + 0.114B \quad I = 0.211R - 0.523G + 0.312B \quad Q = 0.596R - 0.275G - 0.322B$$

由于JPEG只支持YUV颜色模式的数据结构，而不支持RGB图像数据结构，所以在将彩色图像进行压缩之前，必须先对颜色模式进行数据转换。

#### ✧ JPEG算法主要概括：

包括**两个部分**：无损压缩，基于DCT和哈夫曼编码的有损压缩

定义两种压缩算法：基于DCT 有失真算法；基于DPCM 无失真算法

主要存储颜色变化，尤其是亮度变化

与颜色空间无关，算法不包含“RGB 到YUV” 和“YUV 到RGB” 变换

#### ✧ JPEG基于DCT的有损编码算法步骤：

1. **分块：**按序分块，将图像分成 $8 \times 8$  的图像子块为了压缩RGB 彩色图像，这项工作必须进行三遍，分别处理每个颜色，JPEG将原图像划分为若干个子块，每个子块包含8像素 $\times$ 8像素

1. **DCT变换：**对每一个子块，进行DCT 变换，得到DCT 系数矩阵



**3. 量化：**DCT变换的作用是使空间域的能量重新分布，降低图像的相关性。DCT变换本身并不能达到数据压缩的作用，而要实现图像压缩，就要选择适当的比特分配方案和量化方法。量化的作用是在保证主观图像质量的前提下，丢掉那些对视觉效果影响不大的信息。量化是一种降低精度的过程，所以是有损的。会产生一定的失真，失真程度视量化等级而定。

在JPEG标准中采用线性均匀量化器

- 量化过程为对64（8×8）个DCT系数除以量化步长并取整，量化步长由量化表（量化矩阵）决定。量化的计算公式十分简单：  
量化值（ $i, j$ ）=  $T(i, j) / \text{量化矩阵}(i, j)$
- 在解码过程中，逆量化公式为： $T(i, j) = \text{量化值}(i, j) \times \text{量化矩阵}(i, j)$

#### 4. DCT系数编码

（1）直流系数（DC）编码（2）交流系数（AC）编码

**5. 熵编码：**为了进一步达到压缩数据的目的，需要对DPCM编码后的直流系数（DC）和行程编码后的交流系数（AC）再做基于统计特性的熵编码（Entropy Coding）。在JPEG有损压缩算法中，使用哈夫曼（Huffman）编码。哈夫曼编码可以使用很简单的查表（Lookup Table）方法进行编码。在压缩数据符号时，哈夫曼编码对出现频度比较高的符号分配比较

短的代码，而对出现频度较低的符号分配比较长的代码。

✧ JPEG 2000简介（P55）

算法：

JPEG 以离散余弦变换为主的区块编码

JPEG 2000 以离散小波变换算法为主的多解析编码

JPEG2000特点：

（1）高压缩率（2）无损压缩（3）渐进压缩（4）感兴趣区域压缩（5）色彩（6）图像处理简单

#### 3.4.3 视频压缩标准(P58)

**1、MPEG视频压缩算法：**在每帧内利用空间相关性进行压缩，同时在帧之间利用时间相关性进一步压缩。运动图像是静止图像的序列

**主要技术：**（1）减少时间冗余的基于块的帧间运动补偿技术（2）基于DCT变换的减少空间冗余的技术。

**去时间冗余信息** 使用帧间编码技术可去除时域冗余信息，它包括以下三部分：

**A. 运动补偿** 运动补偿是通过先前的局部图像来预测、补偿当前的局部图像，它是减少帧序列冗余信息的有效方法。

**B. 运动表示** 不同区域的图像需要使用不同的运动矢量来描述运动信息。运动矢量通过熵编码进行压缩。

**C. 运动估计** 运动估计是从视频序列中抽取运动信息的一整套技术。

注：通用的压缩标准都使用基于块的运动估计和运动补偿。

**去空间冗余信息** 主要使用帧内编码技术和熵编码技术：

**A. 变换编码** 帧内图像和预测差分信号都有很高的空域冗余信息。变换编码将空域信号变换到另一正交矢量空间，使其相关性下降，数据冗余度减小。

**B. 量化编码** 经过变换编码后，产生一批变换系数，对这些系数进行量化，使编码器的输出达到一定的位率。这一过程导致精度的降低。

**C. 熵编码** 熵编码是无损编码。它对变换、量化后得到的系数和运动信息，进行进一步的压缩。

#### 2. 视频编码

MPEG-1视频压缩技术基本方法可以归纳成两个要点：

A. 在时间方向上，图像数据压缩采用运动补偿（Motion Compensation）算法来去掉冗余信息。

B. 在空间方向上，图像数据压缩采用JPEG（Joint Photographic Experts Group）压缩算法来去掉冗余信息。

**Mp3指的是什么？** MPEG Audio Layer 3 （MPEG-3） MPEG:Motion Picture Experts Group

MP3的音频压缩基于这样一种考虑：因为人耳只能听到一定频段内的声音，而其他更高或更低频率的声音对人耳是没有用处的，所以MP3技术就把这部分声音去掉了，从而使得文件体积大为缩小，但在人耳听起来，却并没有什么失真

## 第四章 多媒体硬件环境

多媒体计算机系统：(1) 多媒体硬件设备 技术要求：

- 主机处理性能强，最好采用高频率、超线程、多内核、低发热的CPU产品
- 主机接口齐全，目前流行的是USB接口
- 各种多媒体设备齐全，硬盘足够大，显示器的分辨率要高。

(2) 多媒体软件

多媒体系统软件，包括操作系统、驱动程序、多媒体开发工具等。

多媒体应用软件，如影音播放软件、教学软件、网络电视软件等。

### 4.1 多媒体存储设备

多媒体存储技术：光存储技术和闪存技术

光存储技术是用激光照射介质，通过激光与介质的相互作用使介质发生物理、化学变化，将信息存储下来的技术。

基本原理：存储介质受到激光照射后，介质的某种性质(如反射率、反射光极化方向等)发生改变，介质性质的不同状态映射为不同的存储数据，存储数据的读出则通过识别存储单元性质的变化来实现。

CD光盘、DVD光盘等利用光存储介质，以二进制数据的形式来存储信息：

- ❖ 信息写入过程中，将编码后的数据送入光调制器，使激光源输出强度不同的光束。调制后的激光束通过光路系统，经物镜聚焦照射到介质上。存储介质经激光照射后被烧蚀出小凹坑，所以在存储介质上存在被烧蚀和未烧蚀两种不同的状态，分别对应两种不同的二进制状态0 或1。
- ❖ 读取信息时，激光扫描介质，在凹坑处入射光不返回，无凹坑处入射光大部分返回。根据光束反射能力的不同，将存储介质上的二进制信息读出，再将这些二进制代码解码为原始信息。

#### 4.1.1 光存储的类型：

- ✧ 只读型光存储系统：只读型光盘包括LV和CD-ROM (Compact Disc-Read Only Memory)等。用户只能从CD-ROM读取信息，而不能往盘上写信息。CD-ROM中的内容在光盘生成时就已经决定，而且不可改变。
- ✧ 一次写型光存储系统：一次写(Write Once Read Many, WORM)光存储系统可一次写入多媒体信息，任意多次读出，与CD-ROM相比，它具有由用户自己确定记录内容的优点。
- ✧ 可重写型光存储系统 可重写型光存储系统(Rewritable 或Erasable, E-R/W)像磁盘一样可任意读写数据。它又分磁光型(Magnetic Optical, MO)和相变型(Phase Change, PC)两种形式

#### 4.1.2 CD-ROM光存储系统P72 (1) CD-ROM盘片 (光盘) (2) 只读光盘驱动器



图 4.1 CD-ROM 的物理层次

光盘驱动器组成：光学头、读通道、聚焦伺服、跟踪伺服、主轴电机伺服和微处理器

CD-ROM 驱动器的工作原理

从CD-ROM光头射出来的激光照到盘片平的地方和小坑的地方反射率不同，这时在激光头旁边的光敏元件，感应到强弱不同的反射光，就产生高低电平，输出到光驱的数字电路，而高低电平在计算机中分别用0, 1表示，这就是CD-ROM把数据光盘转换成数据输出的原理和过程。

#### 4.1.3 CD-R(Compact Disc Recordable)光存储系统 (P73)

一次写，多次读的可刻录光盘系统

- 可重写式光存储系统

➤ **磁光 (MO) 存储系统**：磁光盘片用树脂做基盘，其上集积了保护层（氮化硅）、记录层（**铽、铁钴合金**）和反射层（铝合金）而构成。数据记录在记录层上。形成记录的磁粒子（小磁铁）相对于记录而形成垂直排列（垂直磁记录方式）。由于磁粒子非常细长，垂直排列可以获得较高的记录密度。磁粒子在外磁场的作用下可具有一定的方向性，这种方向性在激光束的热力作用下可发生转，方向性的正、负分别代表0或1，0或1的翻转构成其可擦除性，从擦写次数可达百万次。

➤ **相变 (PD) 光存储系统**：在基盘上沉积电介质层、相变记录层、冷却层和保护层等形成多层结构。

**PD光存储系统读、写和擦除原理**

- 利用物质的状态变化即所谓的相变进行数据的读、写和擦除。
- 相变记录层由一种银合金材料组成，视其加热温度的不同，它可以形成晶体，也可以形成非晶体。
- 适当调整加热温度可以自由地控制记录层的结晶状态。在晶体状态中原子整齐排列，光反射率高；相反，在非晶体状态原子排列不整齐，光的反射率低。

#### 4.1.4 DVD光存储系统 (P74)

**追加写入**是指一次刻录没有刻满，还有可以继续，直至刻满，但是不能删除光盘上的数据，也不能格式化光盘。

##### ◆ DVD特点：

1. 高密度：DVD盘与CD光盘直径均为120mm，但CD光盘仅能存放70分钟视频图像，而单面单层DVD记录层具有4.7GB容量，双面双层光盘的容量高达17GB，可以容纳4部电影于一张光盘上。
2. 高画质：采用国际通用的活动图像压缩标准MPEG 2，可达到广播级的电视质量。
3. 高音质：DVD具有8（7.1）个独立的音频码流，足以实现数字环绕三维高保真音响效果。
4. 高兼容性：DVD视盘机、DVD唱机和DVD-ROM/R/RAM均可播放CD唱盘；DVD视盘机和DVD-ROM/R/RAM均能回放VCD盘；DVD-ROM/R/RAM也可读取CD-ROM盘。
5. 高可靠性：DVD采用RSPC（Reed Solomon Product Code）纠错编码方式和8/16信号调制方式，确保数据读取可靠。

#### 4.1.5 闪存与闪存卡

**非易失随机访问存储器**，俗称闪存(Flash Memory)。特点：断电后数据不消失

**NOR型**：有独立的地址线 and 数据线，成本高，容量小，适合频繁随机读写的场合，通常用于存储程序代码并直接在闪存内运行，如手机。

**NAND型**：地址线 and 数据线是共用的I/O线，成本低，容量大。主要用于存储资料，如闪存卡。

**移动盘**主要包括两类，第一类是由半导体存储器为存储内核，这一类主要是**U盘**。另外一类是以硬磁盘为存储内核，通常称这一类为**移动硬盘**。两类移动存储器都使用计算机的通用接口，目前使用的接口主要是USB接口，但移动硬盘也有使用其它的接口的，如并行口和IEEE-1394。

#### 4.2 音频接口 (P76)

#### 4.3 视频接口 (P82) 显卡（显示适配器）、显示器（CRT、LCD等）、视频卡

##### 新型显示技术：

**等离子显示器**（PDP (Plasma Display Panel)

**液晶显示器**（LCD, Liquid Crystal Display）

**优点**：1、超大平坦的有效显示区；2、采用阵列显示技术，节能效果好；3、提高亮度和对比度，有效减小眩光和反光，具有清晰平滑均匀的显示效果。

**问题**：彩色电视机与彩色显示器的区别？

##### 液晶显示系统：

**LCD构成**：包括背光灯管、导光板、偏光板、玻璃基板、薄膜式晶体管、配向膜、液晶材料、滤光板等等。

**LCD成像过程**：1) 首先LCD利用背光源CCFL萤光灯管投射出光源2) 这些光源会先经过导光板使光线均匀分布在整个屏幕上。3) 然后光线通过一个偏光板，再经过液晶。4) 液晶分子的排列方式会随着控制电压的不同而发生变化，进而改变穿透液晶的光线角度。5) 所有的光线再经过前方的彩色滤光膜与另一块偏光板，才能呈现出各种颜色。6) 因此我们只要改变刺激液晶的电压值就可以控制最后出现的光线强度与色彩，从而能在液晶面板上变

化出有不同深浅的颜色组合了。

屏幕上的图像就象是广告灯箱那样，灯箱里的灯管发出的光透过有图案的薄膜进入人眼。

**LCD电视与CRT电视相比的优点：**超薄、超轻、能耗低、无辐射、不受外磁影响、图像柔和等优点。

**LED是什么？** Light Emitting Diode, 发光二极管的英文缩写，简称LED

LED屏幕通过控制半导体发光二极管的显示方式，用来显示文字、图形、图像、动画、行情、视频、录像信号等各种信息。

用LED灯做为液晶电视的背光源，LED背光源液晶电视成像原理和普通的LCD液晶电视一样，只是将LCD用的CCFL普通背光灯用LED发光二极管代替。

✧ LED背光源与CCFL（LCD）背光源在结构上基本是一致的，其中主要的区别在于LED是**点光源**，而CCFL（LCD）是**线光源**

**为什么要用LED背光源？** 1、**色彩表现力：**由于采用了LED独立发光元件，电视的色域范围大都能达到NTSC 的120%以上，经良好调教可以达到150%。

2、**对比度：**电视可以支持背光区域调整技术，亮度调节很容易实现，因此可达到千万:1级的动态对比度。

#### 4.3.2 视频卡/盒（P87）

#### 4.4 多媒体I/O设备（P88）

##### ✧ 4.4.2 触摸屏（P90）

A. 电阻式触摸屏1）四线电阻式触摸屏2）五线电阻式触摸屏3）六线电阻式触摸屏4）七线电阻式触摸屏 B. 电容式触摸屏1）单点触摸屏2）多点触摸屏（iPhone）C. 红外线触摸屏 D. 表面声波触摸屏 E. 压感触摸屏 F. 电磁感应触摸屏

**电阻式触摸屏：**当手指触摸屏幕时，两层导电层在触摸点位置就有了一个接触，控制器侦查到这个接触，并计算出x, y的坐标。通过接触点电阻的大小可以量化压力大小。

**电容式触摸屏：**当用户触摸电容屏时，由于人体电场，用户手指头和工作面形成一个耦合电容，因为工作面上接有高频信号，于是手指头吸收走一个很小的电流。这个电流分从触摸屏四个角上的电极中流出，并且理论上流经这四个电极的电流与手指到四角的距离成比例，控制器通过对这四个电流比例的精密计算，得出触摸点的位置。

**新式电容触摸屏：**新式电容触摸屏是从电容式触摸按键经过插值算法引申出来的一种触摸屏

检测方法，可以支持多点触摸。如iPhone使用的就是典型的电容触摸感应实现多点触摸。

**红外线触摸屏：**红外触摸屏是在紧贴屏幕前密布X、Y方向上的红外线矩阵，通过不停的扫描是否有红外线被物体阻挡检测并定位用户的触摸。红外线触摸屏可以实现多点触摸检测。

**表面声波触摸屏：**表面声波触摸屏的触摸屏部分可以是一块平面、球面或是柱面的玻璃平板，安装在CRT、LED、LCD或是等离子显示器屏幕的前面。玻璃屏的左上角和右下角各固定了竖直和水平方向的超声波发射换能器，右上角则固定了两个相应的超声波接收换能器。玻璃屏的四个周边则刻有45°角由疏到密间隔非常精密的反射条纹。

##### ✧ 4.4.3 扫描仪

● **扫描仪的工作原理：**扫描图像时，光线从物体上反射回来，通过透镜射进CCD，将光线转换成模拟电压信号，并且标出每个像素的灰度级，再由ADC将模拟信号转换成数字信号，每种颜色使用8，10，12位来表示，扫描后保存在电脑里。

##### ● 扫描仪的分类

1、按扫描方式分类：手动式扫描仪 平面式扫描仪 滚筒式扫描仪 胶片扫描仪

2、按扫描幅面分类 3、按扫描仪分辨率分类 4、按灰度与彩色分类 5、按反射式或透射式分类

##### ● 扫描部件：

CCD：通过镜头聚焦到CCD阵列将光信号转换成电信号成像。**组成：**光学透镜、反射镜条、电荷耦合器件（CCD）组、模数转换电路等。

CIS（接触式图像传感）：CIS紧贴扫描稿件表面进行接触式扫描而成像。

**原理：**以矩阵式排列的内置发光二极管为光源，发出光线被扫描稿件表面反射回来，



被传感器接收后转化为电信号，经数字处理后形成完整的扫描稿件正面影像。

● **扫描仪性能指标：**

(1) 光学精度：可分横向精度（由CCD点数决定）与纵向精度（由步进电机控制）单位：**dpi**，每英寸长度上扫描图像所含像素点的个数。

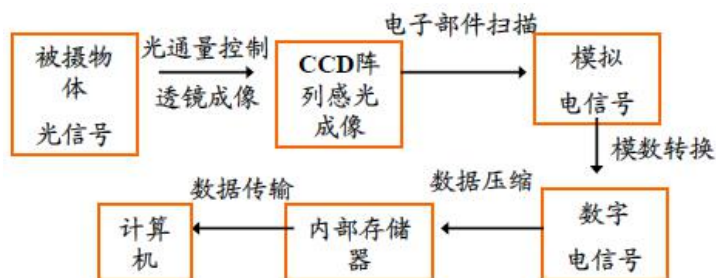
(2) 色彩位数（bit）：扫描仪所能捕获色彩层次信息指标（由模数转换电路决定）

(3) 硬件接口标准：主要有SCSI、EPP（增强打印并口）、USB。

(4) 扫描幅面：幅面越大，价格越高。

❖ **4.4.4 数码相机**

数码相机的工作原理：



数码相机技术指标：

(1) 分辨率：由CCD像素数决定。(2) 色彩浓度（色彩位）。(3) 存储介质。(4) 变焦镜头：自动对焦辅助手动对焦(5) 图像存储格式：BMP（容量大画质好）、JPEG（可调整图像压缩比例）、TIFF（无损压缩或可逆压缩）。(6) 接口标准：串口，SCSI或USB接口。(7) LCD显示屏：2.5英寸

**投影仪：液晶投影机：LCD技术：标准液晶显示技术。**高亮度外光源产生近似纯白光，通过分光、滤光系统分成R、G、B三原色光束，通过三片LCD液晶体。输入信号经处理后分成R、G、B三色驱动信号，分别驱动三片LCD，重现三色图像，经光学系统合成后投射形成还原的彩色图像

投影仪技术指标：

(1) **亮度**：标准测定环境下测定投影的亮度值。(2) **分辨率**：投影机在接收R、G、B视频信号时可达到的最高像素数。与水平/垂直扫描频率、视频带宽有关。(3) **对比度**：反映画面最高与最暗区域亮度之比，从白到黑的渐变层次。对比度越大，投影画面颜色层次越丰富，图像越清晰。(4) **均匀度**：反映边缘亮度与中心亮度的比值。均匀度越高，画面均匀一致性越好。(5) **灯泡寿命（光源）**：影响投影机使用成本。



---

## 第五章 超媒体与Web系统

人类记忆的互联网状结构就和能有多种路径，不同的联想检索必然导致不同的路径

这种互联网状结构用纯文本是无法管理的，必须采用一种比文本更高一层次的信息管理技术，即**超级文本**（Hypertext），超文本与多媒体的融合产生了**超媒体**（Hypermedia）

创作和管理超媒体的系统称为**超媒体系统**

**超媒体与超文本的不同之处：**超文本主要是以文字的形式表示信息，建立的链接关系主要是文句之间的链接关系。

**超媒体**除了使用文本外，还使用声音、图形、图像、动画和视频片段等多种媒体信息来表示信息，建立的链接关系是文本、声音、图形、图像、动画和视频片段之间的链接关系。

**超媒体系统的基本特性体现在：**① 超媒体的数据库是由“声、文、图”类节点或内容组合的节点组成的网络，内容具有多媒体化，网状的信息结构使它的信息表达接近现实世界。

② 屏幕中的窗口和数据库中的节点具有对应关系。③ 超媒体的设计者可以很容易地按需要创建节点，删除节点，编辑节点等，同样也可生成链，完成链接，删除链接，改变链的属性等操作。④ 用户可对超媒体进行浏览和查询。⑤ 具备良好的扩充功能，接受不断更新的超媒体管理和查询技术。

### ● 5.1.2 超媒体的组成要素

**1. 节点（Node）**超媒体是由节点和链构成的信息网络。节点是表达信息的单位，是围绕一个特殊主题组织起来的数据集合。节点的内容可是文本、图形、图像、动画、音频、视频等，也可以是一般计算机程序。

**节点的基本类型**归纳如下：① 文本节点② 图形节点③ 图像节点④ 音频节点⑤ 视频节点⑥ 混合媒体节点⑦ 按钮节点⑧ 组织型节点⑨ 推理型节点

前六种节点类型值表示信息而不表示知识，它们为读、看、听而设计，在面向对象的超媒体系统中，常用到组织型节点和推理型节点

**组织型节点：**包括索引文本节点和索引节点，增加索引是描述节点的一种方法，同时也是数据库管理的需要。

**推理型节点：**是用于辅助链的推理与计算，它包括对象节点和规则节点。可以表现知识的结构和用于保存规则，并且指明满足规则的对象，制定规则是否被使用，进行规则的解释等。

### **1. 链（Link）**

超媒体链又称为**超链**，是节点间的信息联系，它以某种形式将一个节点与其它节点连接起来。

**链的一般结构**可分为三个部分：**链源、链宿及链的属性**。

链源是导致浏览过程中节点迁移的原因，可以是热标、媒体对象或节点等。链宿是链的目的所在，可以是节点，也可以是其他任何媒体的内容。链的属性决定链的基本类型

**各类链的特点**① 基本结构链是构成超媒体的主要形式，在建立超媒体系统前需创建基本结构链。它的特点是层次与分支明确。② 索引链是超文本所特有的③ 推理链用于系统的机器推理与程序化。④ 隐形链又称关键字链或查询链。

**3. 热标（Hotspot）**确定信息之间的关联，引起向相关节点的转移。不同的媒体有不同的热标 **1) 热字（hot-word）**

**2) 热区（hot-area）：**在所显示的图像的显示区上指明一个敏感

区域作为触发转移的源点**3) 热元（hot-element）：**在图形媒体中，以相对独立的图形基本单位作为信息转移的链源，如一个圆形，一条线等**4) 热点（hot-point）：**主要用于时基类媒体，如视频，声音等在时间轴上的触发转移**5) 热属性（hot-attribution）：**把关系数据库中的属性作为热源使用

**4. 宏节点** 宏节点是指链接在一起的节点群，更准确地说，一个宏节点就是超文本网络的一个有某种共同特征的字集。

➤ 超文本与超媒体系统就是完成生成并只使用前述超文本或超媒体的软硬件的总称。

从理论上讲可将其划分为三个层次：1. 表现层——用户接口；2. 超文本抽象机层——节点和链；3. 数据库层——存储、共享数据和网络访问

**超文本与超媒体系统的两个模型：**

超文本和超媒体的系统结构较著名的是Campbell和Goodman模型，

---

另一个是从事超文本标准化研究Dexter小组提出的Dexter模型。

## 5.2 Web超媒体系统

World Wide Web 简称WWW或W3，译为万维网。

Web 超文本系统3层结构：1) 表现层2) 超文本抽象机器层3) 超文本信息库层

支持WWW客户/服务器的协议主要有两个：

**超文本传输协议 (Hyper Text Transfer Protocol , HTTP)**。它为客户/服务器通信提供了联络方式及信息传送格式。

**超文本标记语言 (Hyper Text Markup Language , HTML )**。它是一种用户与程序都能理解的语言，它是为文献提供表现界面与超文本链接的标记语言。

**HTTP与WEB服务：**1. HTTP应用层协议是Web的核心。2. HTTP协议运行在客户程序和服务器程序中。不同端系统上的客户程序和服务器程序通过交换HTTP消息彼此交流。3. HTTP定义Web客户(即浏览器)如何从web服务器请求Web页面，以及服务器如何把Web页面传送给客户。

HTTP定义的事务处理由以下四步组成：1. 客户端与服务器端建立连接2. 客户端向服务器端发送请求 3. 服务器端向客户端回复响应4. 断开连接

➤ **HTTP的特点：客户/服务器模式：**一个服务器可以为分布在世界各地的许多客户服务。**简单：**HTTP本身处理简单，有效地处理大量请求，HTTP服务器程序规模小，所以经由HTTP的通信速度快，与其它协议相比，时间开销小得多。**灵活：**HTTP允许传输任意类型的数据对象，可以通过Content-type来指定数据类型。**无状态：**HTTP是无状态的协议，缺少状态记忆，运行速度高，服务器应答速度较快。

➤ **HTTP的消息类型共分为两大类 (P156)**

**请求 (Request) 消息：**由客户端发给服务器的消息。其组成包括：请求行 (Request-Line)，可选的头域 (Header Field)，及实体 (Entity-Body)。

**响应 (Response) 消息：**是服务端回复客户端请求的消息，其组成包括状态行 (Status-Line)，可选的头域 (Header Field)，及实体 (Entity-Body)。

**超文本标记语言：SGML、HTML、XML、VRML**

- **SGML标准通用标记语言 (Standard Generalized Markup Language, SGML)** SGML的主要特点是：它不仅支持无数的文档结构类型，如布告、技术手册、章节、目录、设计规范、各种报告、信函和备忘录等，还可以创建与特定的软硬件无关的文档，方便不同计算机系统的用户交换文档。
- **HTML的特点：**HTML文档制作不是很复杂，且功能强大，支持不同数据格式的文件嵌入，这也是WWW盛行的原因之一，其主要特点如下：1 简易性，HTML版本升级采用超集方式，从而更加灵活方便。2 可扩展性，HTML语言的广泛应用带来了加强功能，增加标识符等要求，HTML采取子类元素的方式，为系统扩展带来保证。3 平台无关性。虽然PC机大行其道，但使用MAC等其他机器的大有人在，HTML可以使用在广泛的平台上，这也是WWW盛行的另一个原因。
- **XML可扩展标记语言 (Extensible Markup Language, XML)** 是一种新型的标记语言。它同HTML一样，都来自标准通用标记语言 (Standard Generalized Markup Language, SGML)。早在Web未发明之前，SGML就早已存在。正如它的名称所言，SGML是一种用标记来描述文档资料的通用语言，它包含了一系列的文档类型定义 (简称DTD)，DTD 中定义了标记的含义，因而SGML 的语法是可以扩展的。
- **VRML WWW上的虚拟现实技术是依靠虚拟现实造型语言 (Virtual Reality Modeling Language, VRML)** 来实现的，VRML是一种用来描述万维网页面上三维交互环境的文件格式，它的基本原理同HTML的基本原理一样简单，都是用一系列指令告诉浏览器如何显示一个文档，它们都是描述万维网页面的描述语言。与HTML不同的是，以HTML为核心的万维网浏览器浏览的是二维世界，而以VRML为核心的万维网浏览器浏览的是三维世界，你可以使用鼠标在这个世界里到处“逛一逛”而不是像二维世界里“一页一页”地显示。

**虚拟现实在Internet上的应用有：**远程教育 商业应用 娱乐

### 超文本与超媒体存在的问题

超文本与超媒体是一项正在发展中的技术，虽然它有许多独特的优点，但也存在许多不够完善的方面。

1. **信息组织** 超文本的信息是以节点作为单位。如何把一个复杂的信息系统划分成信息块是一个较困难的问题。
2. **智能化** 虽然大多数超文本系统提供了许多帮助用户阅读的辅助信息和直观表示。但因超文本系统的控制权完全交给了用户，当用户接触一个不熟悉的题目时，可能会在网络中迷失方向。
3. **数据转换** 超文本系统数据的组织与现有的各种数据库文件格式完全不一样。
4. **兼容性** 目前的超文本系统大都是根据用户的要求分别设计的，它们之间没有考虑到兼容性问题，也没有统一的标准可循。所以要尽快制定标准并加强对版本的控制。标准化是超文本系统的一个重要问题，没有标准化，各个超文本系统之间就无法沟通，信息就不能共享。
5. **扩充性** 现有的超文本系统，有待于提高检索和查询速度，增强信息管理结构和组织的灵活性，以便提供方便的系统扩充手段。
6. **媒体间协调性**

**超文本与超媒体发展的前景**• 由超文本向超媒体发展• 由超媒体向智能超媒体发展• 由超媒体向协作超媒体发展

## 第六章 多媒体数据库

**多媒体数据管理的问题**1. 人工管理阶段2. 文件系统阶段3. 数据库系统阶段

**数据库 (DB-Data Base)** 是按照一定的规则相互关联数据的集合。这些数据是用来反映各类实物的属性的数据。

**数据库系统 (DBS-Database System)** 是指具有管理和控制数据库功能的计算机系统。

**数据库管理系统 (DBMS-Data Base Management System):** 是能够用来建立数据库，完成数据库的维护、管理等操作，并能够帮助用户方便地使用数据库的软件。

- **数据库系统由5部分组成:** (1) **硬件系统**是整个数据库系统的基础，需要有足够大容量的内存和磁盘等存取设备等。(2) **数据库集合**是若干个设计合理、满足应用需要的数据库。(3) **数据库管理系统**是为数据库的建立、使用和维护而配置的软件，是数据库系统的核心组成部分。(4) **相关软件**是支持软件，如操作系统等。(5) **数据库管理员**是全面负责建立、维护和管理数据库系统的人员；用户是最终系统的使用和操作人员。

数据库系统的一个重要概念是数据独立性。

**传统的数据库系统**，一般按层次被划分为三种模式：1、物理模式（存储模式）：数据的存储组织方法2、概念模式（逻辑模式）：定义抽象世界的方法 3、外部模式（视图）：概念模式对用户有用的那一部分

❖ **数据库系统的特点:** **数据共享:** 多个用户可同时存取数据而不相互影响； **减少数据冗余:** 在数据库中用户的逻辑数据文件和具体的物理数据文件不必一一对应； **具有较高的数据独立性:** 指数据与应用程序之间的彼此独立 **增强了数据安全性和完整性保护:** 加入安全保密机制，实行集中控制。

**传统数据库数据模型:** 1、关系型 2、层次型 3、网络型

**关系型** 采用关系框架来描述数据之间的关系，通过把数据抽象成不同属性和相互的关系，建立起数据的管理机制。列称为字段，行称为记录。

**多媒体数据库中的数据类型:** 1、字符数据2、文本数据3、声音数据（音频）4、图形数据5、图像数据6、视频数据

**多媒体数据的特点:** (1) 数据量巨大 (2) 数据类型繁多、差别大 (3) 有些多媒体信息对处理时间要求过高 (4)

多媒体信息往往需要多种信息集成、共同描述

❖ **多媒体数据的管理方法：**（1）文件管理系统（2）建立特定的逻辑目录结构（3）传统的字符、数值数据管理系统（4）多媒体数据管理系统 特点：如果能够对多媒体数据和传统的字符型数值数据采用统一的方法进行管理，该系统将成为最理想、最有效的数据管理方法。功能：a. 多媒体数据库对象的定义b. 多媒体数据库运行的控制 c. 多媒体数据库的建立和维护d. 多媒体数据库在网络上的通信功能 主流数据库系统：如Oracle、Sybase 等（5）**超文本和超媒体**。超文本：其中的某些字、符号或短语起着“热链路”（Hotlink）的作用，采用一种非线性的网状结构组织块信息。超媒体：超文本可利用“热链路”链接其他不同类型（内含声音、图片、动画）的文件，这些具有多媒体操作的超文本，称为超媒体。

**多媒体数据库管理系统的基本功能：**（1）必须能表示和处理各种媒体数据（包括音频、视频、动画等非格式化数据）。（2）必须能反映和管理各种多媒体数据的特性，或各种媒体数据之间的空间和时间的关联。（3）必须满足物理数据独立性和逻辑数据独立性，同时满足媒体数据独立性。 物理数据独立性：物理数据结构（存储模式）改变不影响概念数据组织（逻辑模式）。 逻辑数据独立性：概念数据组织改变时，不会影响用户程序使用的视图（外模式）。 媒体数据独立性：多媒体数据库管理系统的设计和实现时，要求系统保持各种媒体的独立性和透明性。

➤ **6.2.1 多媒体数据库管理系统的体系结构** 1、层次结构（P175）2、组织结构（P173 联邦型结构..）

➤ **6.3 多媒体数据模型（P176）**

**1. NF2数据模型** 所有的关系数据库中的关系必须满足最低的要求。这个要求就是**第一范式**，简称1NF。表中不能有表。由于多媒体数据库中具有各种各样的媒体数据，这些媒体数据又要统一地在关系表中加以表现和处理，就不能不打破关系数据库中关于范式的要求，要允许表中有表，这就是NF2(Non First Normal Form) 但是其**局限性很大，建模能力不强**。

1NF(Normal Form)：第一范式，将关系规定成二维表格形式 NF 2 (Nor First Normal Form):非第一范式

**2、面向对象数据模型：**（1）对象：问题领域中事物的表示和描述（2）属性：组成对象的数据（3）方法：定义在对象属性上的一组操作（4）消息：对象间的通信和请求对象完成某种处理工作要传递的信息（5）类：具有相似属性的一组对象（6）基类和派生类（7）继承性：建造新的派生类（从个别基类继承属性和方法，并增加新定义的属性和方法）（8）封装性：对象的实现细节对外部是不透明的（不可见的），也是不可取的。（9）多态性：同样的消息引起不同对象的不同操作

**面向对象数据模型（特点）：**（1）以客观自然的方法来描述现实世界中的各种实体及实体间的联系（2）更具表示能力（支持抽象数据类型和用户定义方法）（3）封装性使得多媒体对象的界面独立于对象的内容表示（4）基类、派生类及其继承性可减少冗余和避免引起问题，利于版本控制（5）查询语言是沿着系统内部固有联系进行，避免大量的查询优化工作

**3、其他数据模型** 超媒体数据模型；文献模型； 专有媒体数据模型；

## 第七章 网络多媒体技术

**分布式多媒体应用的通信需求。**多媒体通信是通信技术和多媒体技术相结合的产物，具有计算机的交互性、多媒体的复合性、通信的分布性及电视的真实性。

**1. 多媒体数据流的基本特征** (1) **比特率可变性** 恒定比特率(CBR) 可变比特率(VBR) (2) **时间依赖性** 传输实时，端到端等待时间应控制在一个很短的时间段内 (3) **信道对称性** 上行信道和下行信道 对称和不对称 (VOD, 对等式视频会议)

**2. 多媒体通信的性能需求** (1) 吞吐量需求 (2) 可靠性需求 (3) 延迟需求 (4) 多点通信需求 (5) 同步需求

- **吞吐量需求：**网络吞吐量是指有效的网络带宽，通常定义为物理链路传输速率减去各种传输开销(物理传输开销、网络冲突、拥塞和差错等开销)。影响吞吐量的因素：网络故障、网络拥塞、瓶颈、缓冲区容量和流量控制等。多媒体通信要求高传输带宽、大缓冲容量
- **可靠性需求：**网络差错主要由位出错、分组丢失和乱序等原因引起的。**差错率**反映了网络传输的可靠性，可用以下3种方法定义：位差错率(BER)—物理传输网，例如SONET 帧差错率(FER)—ATM网 分组差错率(PER)—分组交换网

多媒体应用允许网络传输中存在一定的错误—>人类感知能力的限制。音频比视频的可靠性需求要高一些。

- **延迟需求：****端到端延迟**是指发送端发送一个分组到接收端正确地接收该分组所经历的时间。主要包括：(1) **传播延迟：**传输一个二进制位，为常数；(2) **传输延迟：**传输一个数据块，与网络传输速率和中间节点处理延迟有关；(3) **网络延迟：**传播延迟和传输延迟之和；(4) **接口延迟：**发送端从开始准备发送数据块到实际利用网络发送需要的时间。**延迟抖动：**在一条连接上分组延迟的最大变化量，即端到端延迟的最大值和最小值之差。
- **多点通信需求：**通信种类： 1. 点到点通信； 2. 多播通信(组播)：把相同的数据传送给其它相关站点； 3. 广播通信：把相同的数据传送给其它所有站点。
- **同步需求：**多媒体通信的同步有两种：**流内同步：**保持单个媒体流内部的时间关系，以满足感官上的需求；否则，音频断续，视频不连续 **流间同步：**不同媒体间的同步。需要在目的地对这些媒体流进行同步。
- ◆ **服务质量：(Quality of Service, QoS)**用于说明网络“好坏”的程度。多媒体网络系统必须提供QoS参数定义和相应的管理机制。QoS={吞吐量, 差错率, 端到端延迟, 延迟抖动}

### 多媒体通信网络环境

**计算机网络的硬件组成** 1. **服务器**(网络操作系统，提供共享服务) 2. **工作站** 3. **网卡** 4. **连接介质：**无线、有线(双绞线、同轴电缆、光纤、电话线) 5. **连接设备：**集线器(HUB)、交换机(Switch)、路由器(Router)、网桥(Bridge)、中继器(Repeater)、网关、调制解调器(Modem)等

### ❖ OSI RM: Open System Interconnection Reference Model 开放系统互连参考模型

- **物理层：**OSI模型的最底层。它提出了网络的物理特性，比如连接的电缆类型。这里是二进制值0和1的世界，也就是数据以信号的电特性(高低电平)来表示。**网卡、集线器、中继器、传输介质连接器、编码-解码器**等都工作在物理层。主要讨论在通信线路上比特流的传输问题。这一层协议描述传输媒质的电气、机械、功能和过程的特性。其典型的设计问题有：信号的发送电平、码元宽度、线路码型、物理连接器插脚的数量、插脚的功能、物理拓扑结构、物理连接的建立和终止、传输方式等
- **数据链路层：**指明将要发送的每个数据包的大小、每个数据包的地址以使它们送到指定的接收者那里。也能提供基本的错误识别和校正机制，以确保发送的数据和接收的数据一样。**交换机**工作在OSI参考模型的第二层——数据链路层上，主要功能包括物理编址、网络拓扑结构、错误校验、帧序列以及流控。
- **网络层：**就是告诉数据包从一个网络到另一个网络怎样走(术语叫“路由”)。**路由器**工作在OSI模型中的第三层，即网络层。主要处理分组在网络中的传输。这一层协议的功能是：路由选择、数据交换，网络连接的建立和终止，一个给定的数据链路上网络连接的复用，根据从数据链路层来的错误报告而进行的错误检测和恢复，分组的排序，信息流的控制等。**网络层功能：**1. 定义与指定协议相关联的源和目标逻辑地址 2. 定义通过网络的路径 3. 多链路连接 **什么是路由？**路由是指导IP报文发送的路径信息



- **传输层**：通过一个唯一的地址指明计算机网络上的每个节点，并管理节点之间的连接。同时将大的信息分成小块信息，并在接收节点将信息重新组合起来。**传输层功能**：1. 区分不同的上层应用（分段）2. 建立应用间的端到端连接 3. 定义流量控制（缓存技术、源抑制报文、窗口机制） 4. 为数据传输提供可靠或不可靠的连接服务
- **会话层**：在网络节点之间建立“会话”。**会话层的主要功能**：在建立会话时核实双方身份是否有权参加会话；确定何方支付通信费用；双方在各种选择功能方面取得一致；在会话建立以后，需要对进程间的对话进行管理与控制，例如对话过程中某个环节出了故障。
- **表示层**：负责把网络上传输的数据从一种陈述类型转换到另一种类型，也能在数据传输前将其打乱，并在接收端将其恢复。**表示层**主要处理应用实体间交换数据的语法，其目的是解决格式和数据表示的差别，从而为应用层提供一个一致的数据格式，如文本压缩、数据加密、字符编码的转换，从而使字符、格式等有差异的设备之间相互通信。
- **应用层**：OSI的最高层，讨论应用程序用于同网络通信所需要的技术。在这里，我们可以看到很多熟面孔，比如HTTP（超文本传输协议），FTP（文件传输协议），WAP（无线应用协议），SMTP（简单邮件协议）等等。应用层与提供网络服务相关，这些服务包括文件传送、打印服务、数据库服务、电子邮件等。应用层提供了一个应用网络通信的接口

**路由器：运行在网络层** 广播信息控制；多点发送信息控制；路径优化；流量管制；逻辑寻址；提供WAN连接

#### ✧ 路由与交换的区别

- 交换机在二层处理数据包，速度快，但容易造成广播风暴
- 路由器是第三层设备，很早就作为网络连接的设备出现了。路由器处理包速度慢，但在隔阻广播风暴方面相当有效

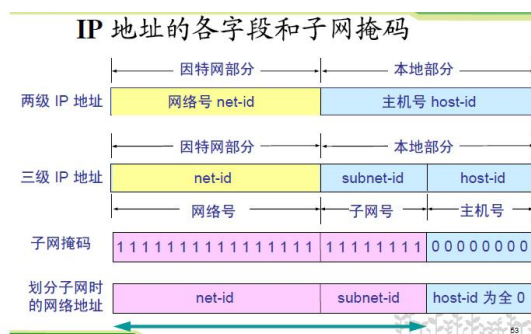
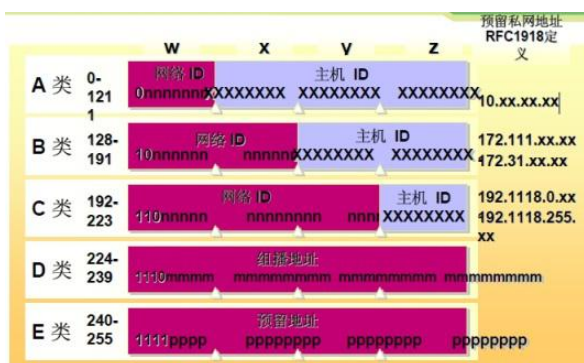
#### ❖ IP地址的编址方法：分类的IP地址；子网的划分

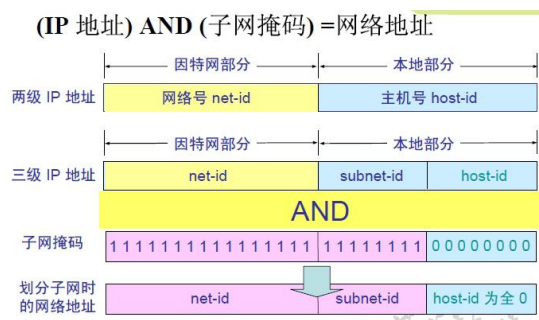
##### ➤ 分类的IP地址

• 每一类地址都由两个固定长度的字段组成，其中一个字段是**网络号net-id**，它标志主机（或路由器）所连接到的网络，而另一个字段则是**主机号host-id**，它标志该主机（或路由器）。

• 两级的IP地址可以记为：

IP地址 ::= { <网络号>, <主机号> } ::= 代表“定义为”





A 类、B 类和 C 类 IP 地址的默认子网掩码		
A 类地址	网络地址	net-id   host-id 为全 0
	默认子网掩码 255.0.0.0	11111111   000000000000000000000000
B 类地址	网络地址	net-id   host-id 为全 0
	默认子网掩码 255.255.0.0	1111111111111111   0000000000000000
C 类地址	网络地址	net-id   host-id 为全 0
	默认子网掩码 255.255.255.0	11111111111111111111   00000000

## ➤ 子网的划分:

- 从1985 年起在IP 地址中又增加了一个“子网号字段”，使两级的IP 地址变成成为三级的IP地址。
- 这种做法叫作划分子网(subnetting)。划分子网已成为因特网的正式标准协议。
- 划分子网纯属一个单位内部的事情。单位对外仍然表现为没有划分子网的网络。
- 从主机号借用若干个比特作为子网号subnet-id，而主机号host-id 也就相应减少了若干个比特。

IP地址::={<网络号>,<子网号>,<主机号>}

- 当没有划分子网时，IP 地址是两级结构，地址的网络号字段也就是IP 地址的“因特网部分”，而主机号字段是IP 地址的“本地部分”。
- 划分子网后IP 地址就变成了三级结构。划分子网只是将IP 地址的本地部分进行再划分，而不改变IP 地址的因特网部分。
- 从一个IP数据报的首部并无法判断源主机或目的主机所连接的网络是否进行了子网的划分。
- 使用子网掩码(subnet mask)可以找出IP地址中的子网部分。

局域网 (Local Area Network, LAN) 局域网是在一个较小的区域内的网络，通常是指一幢楼内或一个单位内。

广域网 (Wide Area Network, WAN) 又称远程网，它的作用范围通常为几十到几千公里，甚至整个世界。

ATM网络 ATM网络是一种面向连接的交换式网络。

## 7.4 多媒体通信协议

通信协议是指通信双方必须遵循的控制信息交换的规则之集合。

通信协议: 在网络基础结构上提供面向连接或无连接的数据传输服务，以支持各种网络应用。

局域网中传输文件-----NetBEUI协议 在局域网中联机玩游戏-----IPX/SPX协议 通过局域网访问Internet---TCP/IP协议

### 7.4.1 IPv6协议

IP v6的一个重要设计目标是保证与Ip v4兼容。 128位的地址空间允许更多的主机被寻址，并且允许地址层上有更多的层次。 改进的多站点寻址方案允许将多站点路由限制在指定的范围内。 组块头的新定义的“流标志字段”允许鉴别属于同一数据流的所有组块。 用于真实性、完整性及数据加密性的新机制。

#### IPv6 与IPv4的比较

- 相同点: 非连接方式; 每个分组根据其包含的目的地址在网络内独立流动; 两者都有跳数限制; IPv6保留了IPv4各选项所提供的绝大多数设施。
- 不同点:
  - ◎地址域的长度: IPv6为128bits, 而IPv4是32bits, 扩大的地址空间可满足互联网发展的需要。
  - ◎标头格式: IPv6完全不同于IPv4, 域的数目由12 减为8, 几乎每个域都被改变。
  - ◎扩展标头: IPv4只用一个标头, 用选项扩展功能; IPv6一般有多个标头 (包括一个基本标头和0或多个扩展标头), 支持可选的功能扩展。
  - ◎支持视频音频流应用: IPv6有建立高质量通路的机制。

IP v6目前定义了三种地址: 单播(Unicast)、多播(Multicast)和任播(Anycast), 利用地址格式前缀表示各种类型。

## 7.5 流媒体技术 流媒体指在网络中使用流式传输技术传送的连续且与时间相关的媒体

网络上传播多媒体信息主要有两种方式: 下载和流式传输。流媒体是相对于传统的下载一回放方式而言的一种媒体格式 流媒体实现的关键技术就是流式传输。

流式传输定义很广泛，现在主要指通过网络传送媒体（如音频、视频）的技术总称。

实现流式传输有两种方法：顺序流式传输（Progressive Streaming）和实时流式传输（RealTime Streaming）。

### ➤ 流媒体播放方式

**单播：**在客户端与媒体服务器之间需要建立一个单独的数据通道，从一台服务器送出的每个数据包只能传送给一个客户

机，这种传送方式称为单播。

**组播：**IP组播技术构建一种具有组播能力的网络，允许路由器一次将数据包复制到多个通道上。

**点播与广播：**点播连接是客户端与服务器的主动的连接。广播指的是用户被动接收流。

例如：aaaa bbb cc d eeee ffffff

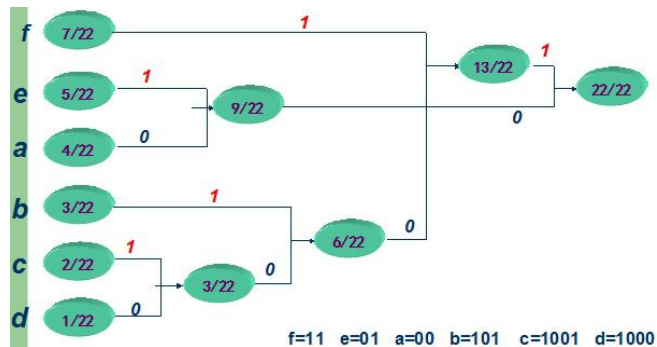
分布为：

a:4/22 b:3/22 c:2/22

d:1/22 e:5/22 f:7/22

排序为：

d c b a e f  
1/22 2/22 3/22 4/22 5/22 7/22

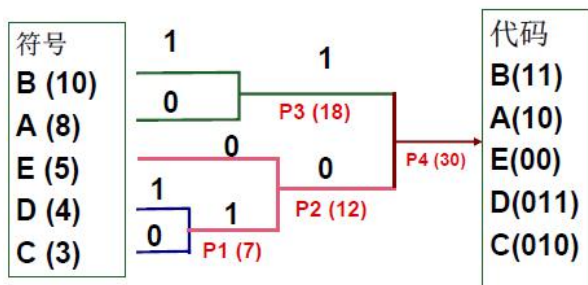


### • 霍夫曼编码举例1

— 现有一个由5个不同符号组成的30个符号的

字符串：**BABACACADADABBCBABEBEDDABEEEBB**

— 计算(1) 该字符串的霍夫曼码(2) 该字符串的熵(3) 该字符串的平均码长(4) 编码前后的压缩比



5个符号的代码

符号	出现的次数	$\log_2(1/p_i)$	分配的代码	需要的位数
B	10	1.585	11	20
A	8	1.907	10	16
C	3	3.322	010	9
D	4	2.907	011	12
E	5	2.585	00	10
合计	30	1.0		67

### (2) 计算该字符串的熵

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

其中， $X = \{x_1, \dots, x_n\}$  是事件 $x_i (i=1, 2, \dots, n)$  的集合，

$$\sum_{i=1}^n p(x_i) = 1$$

并满足

$$- H(S)$$

$$\begin{aligned}
 &= (8/30) \times \log_2(30/8) + (10/30) \times \log_2(30/10) + \\
 &\quad (3/30) \times \log_2(30/3) + (4/30) \times \log_2(30/4) + \\
 &\quad (5/30) \times \log_2(30/5) \\
 &= [30 \lg 30 - (8 \times \lg 8 + 10 \times \lg 10 + 3 \times \lg 3 + 4 \\
 &\quad \times \lg 4 + 5 \lg 5)] / (30 \times \log_2 2) \\
 &= (44.3136 - 24.5592) / 9.0308 = 2.1874 \text{ (Sh)}
 \end{aligned}$$

---

(3) 计算该字符串的平均码长

– 平均码长:  $l = \sum_{i=1}^N l_i p(l_i)$

$$= (2 \times 8 + 2 \times 10 + 3 \times 3 + 3 \times 4 + 2 \times 5) / 30 \\ = 2.233 \text{ 位/符号}$$

平均码长:  $67/30=2.233$ 位

(4) 计算编码前后的压缩比

- 编码前: 5个符号需3位, 30个字符, 需要90位
- 编码后: 共67位

压缩比:  $90/67=1.34:1$