

第一章 多媒体的基本概念

1—1 多媒体——计算机发展的一大趋势

八十年代多媒体技术的崛起和飞速发展，使之成为计算机领域里的一棵奇葩，引人注目。有人把它称之为继纸张印刷术、电报电话、广播电视、计算机之后，人类处理信息手段的一大飞跃，是计算机技术的又一次革命。人们相信，多媒体技术会极大地改变人们的生活方式，推动许多产业的发展，并将导致现有产业结构的调整。最终，很可能计算机、通信、大众传播、娱乐业等会走向大同，形成统一的多媒体产业。

现在，“多媒体”这一术语不仅在计算机界广为流行，而且在其他技术领域和报刊杂志上屡见不鲜；在展览会上，多媒体展品格外引人注目。近几年，在对全世界计算机界最有影响的美国计算机大展 COMDEX 上，有关多媒体的展台约占一半。在这些展台上，图、文、声并茂，琳琅满目，热闹非凡。一次多媒体国际学术会议竟会有数万人蜂拥而至；在被称作多媒体年的 1993 年 Las Vegas 的 COMDEX 上，参展的厂商几乎达 2000 个，有 170000 人出席参观展览，盛况空前……。

一时间，多媒体成了人们关注的热点，人们不约而同地问：什么是多媒体？

大家知道，在人类社会中，信息的表示形式是多种多样的，人们把这些信息的表示形式（或者传播或存储形式）叫做“媒体”。通常，人们见到的文字、声音、图像、图形等都是信息表示的媒体，因此可以说多媒体是“古已有之”，并不稀奇。

为什么近些年才提出“多媒体”呢？一是人们已经有了把多种媒体信息统一处理的需要；而更重要的是，随着技术的发展，人们已经拥有快速处理、存储、传输多媒体信息的能力，这才使“多媒体”变为人类社会生活中的一种客观现实。

这里也应指出，现在当人们在说“多媒体”的时候，常常不完全是在说多媒体信息本身，而主要是指处理和应用它的一套技术。因此，“多媒体”常常就被当作“多媒体技术”的同义语来使用。

此外，还应注意到，现在人们谈论多媒体技术时，常常是站在计算机的立场上来讨论问题。诚然，计算机的数字化技术及交互式的处理能力等，的确为多媒体技术的发展提供了前提条件。但是，也应当认识到，多媒体并非计算机界的“专利”。近年来，音像、影视界的发展也大量采用计算机技术，而且发展很快，因而那些音像、影视系统的人士和生产厂家也会提出他们自己的看法，比如他们可能把计算机界称之为“多媒体计算机”的系统叫作“多媒体电视”等等。这样做没有什么不好，相互间的竞争，从不同的角度，从不同的需要去提出和解决问题，恰恰可能对多媒体技术的日益发展和完善起着极大的推动作用。

现在，我们甚至可以大胆预言，多媒体计算机的进一步发展，很可能使它能在不久的将来占领一切目前拥有计算机、电视机、录像机、音响的领地，和一切文化及娱乐场所，其市场可能是计算机和音像系统的总和，并最终形成一个庞大而完整的多媒体产业。

1—2 信息与媒体

信息与物质和能源一样是当今世界的一项重大资源，信息产业被称作“现代社会的先驱产业”。信息科学技术革命已成为现代科学技术革命的主流。若干发达国家和地区已把信息产业作为第一产业。据估计，到 2000 年，信息产业将成为世界第一大产业。显然，信息在政治、经济、军事、文化及社会进步等各方面都起着至关重要的作用。

那么，什么是信息呢？信息的定义有许多种，据统计大概有三十多种。这里，我们采用下面的定义，即：“信息是通过交流得到的关于特定事物的知识”。这一定义关心的是“内容”，是“知识”本身。这与控制论先驱者维纳：“信息关心的是内容”的观点一致。

另一种定义对从事信息处理的人们或许更有价值，这个定义为：“信息是任何可以通过编码而为计算机处理的数据”。

什么是媒体呢？媒体（Media）是表示、传输、存储信息的载体。

在进一步讨论媒体之前，我们首先了解一下人类感知信息的途径，这些途径有：

- 视觉 是人类感知信息的最重要的途径，约占人类从外部世界获取信息总量的 80%。

其重要性十分明显，以后的许多章节都要重点讨论这方面的内容。

- 听觉 约占人类从外部世界获取信息总量的 10%。

- 嗅、味、触觉 共约占 10%。

下面我们来看看相应的媒体，这里，以视觉、听觉为例：

- 视觉信息

—表示媒体：文字、图画、视频

—存储媒体：纸、录相带、光盘、磁盘

—传输媒体：光、电波

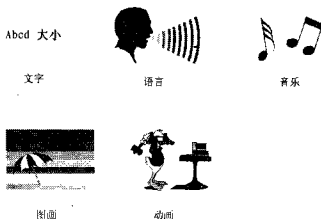


图 1—1 表示媒体示例

• 听觉信息

- 表示媒体：语言、音乐
- 存储媒体：唱片、录音带、光盘、磁盘
- 传输媒体：空气、电话线、电波

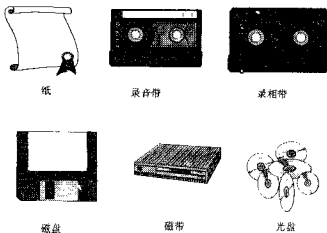


图 1-2 存储媒体示例

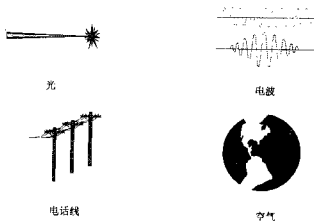


图 1-3 传输媒体示例

1—3 多媒体计算机

1—3—1 多媒体计算机——计算机发展的必然结果

前已述及，之所以出现“多媒体”，一是人们已经有了把多种媒体信息统一处理的需要；而更重要的是，随着技术的发展，已经拥有快速处理、存储、传输多媒体信息的能力。现在，对计算机而言，人们已不再仅仅满足于计算机的速度、容量，也不是主要用它来进行计算，而是用来获取、处理、存储、传输各种需要的信息，而我们现在讨论的信息，实际上就是多媒

体信息。所以，多媒体计算机的出现是计算机发展的一种必然结果。

1—3—2 对多媒体计算机的要求

按照 Apple 公司 Wollaston 的说法，多媒体是文字、图形、动画、视频和音频的结合，而该公司的 Macintosh 机则是把它作为结合起来的“胶水”。Sun Microsystems 也使用了这个定义，但是强调了使用的工具，按该公司 Jeef Morgan 的说法，多媒体是“传统的计算机——文字、图形、图像等与视频、音频、以及为了知识创建和表达的交互式应用的结合体”。实际上，还有人把多媒体看成是三种“革命”因素的总和。其一是在通讯中结合了电视中的音像(Audio-visual)能力；其二是出版发行的能力；其三是计算机交互的处理能力。

现在，在总结、分析的基础上，对多媒体计算机提出了若干要求——亦即对于多媒体计算机，它应能做到：

- 能处理多种感觉(perception)媒体

计算机能处理的感觉媒体有六种：文字、图形、图像（上述三种为静态媒体）、声音、动画、活动影像(video)。（上述三种为时变媒体）

- 在处理的媒体中，至少一种是时变媒体(time-variant)
- 各媒体具有协同性（即为了达到某一目的）
- 交互性

对读者来说，记住上述的全部要求虽说不难，但并不太方便。我们认为，可以给多媒体（实际上是多媒体技术或多媒体计算）下这样一个简单的定义：

多媒体是能够同时抓取、处理、编辑、存储和展示两个以上不同类型信息媒体的技术。这些信息媒体包括：文字、图形、图像、声音、动画、活动影像等。

要强调的是，正是由于利用了计算机中的数字化技术和交互式处理能力，才使多媒体技术成为可能，才能对多种信息媒体进行综合统一的处理。这也就是为什么一般具有声音，图像的电视机、录像机还谈不上是“多媒体”的原因。

1—3—3 MPC 的技术规范标准

MPC 是多媒体计算机的简称。1991 年 11 月，由 Philips, Microsoft 等 14 家有名厂商组成了多媒体市场协会。MPC 是多媒体市场协会制定出来的标准。实际上存在两个标准：第一个层次的 MPC 标准和第二个层次的 MPC 标准。第一个层次的 MPC 标准是 1991 年 11 月提出，建立在 10MHz 的 286AT 的基础之上的（很快又修改成了采用 16MHz 的 386SX）。较为实用的第二个层次的 MPC 标准则是 1993 制定出来的，这一技术规范定义了第二个层次的最小系统功能，但并不打算推荐它作为特定系统的结构。MPC 的标准由多媒体 PC 市场委员会管理（其通讯处：Multimedia PC Marketing council, 1730 M Street NW, Suite 707, Washington, DC 20036 USA）。

一、第一个层次的 MPC 技术规范标准

1. 硬件平台

- 80386SX 以上的 CPU
- 2 MB RAM
- 30 MB 硬盘

- VGA (16 色)
 - CD—ROM 驱动器, 应有:
 - CD 数字音频输出
 - 传输速率不小于 150K/秒, 占用 CPU 的开销应小于 40%
 - 平均寻址时间不超过 1 秒
 - 音频卡
 - 数字化录音 (AD 转换), 8 位量化精度, 采样频率 11.025kHz
 - 放音 (DA 转换), 8 位量化精度, 采样频率为 11.025 或 22.05kHz, 且放音与录音之间的时间误差小于 2%
 - 内置声音合成器
 - 内置调音台
 - 占用 CPU 的开销小于 20%
 - 两键式鼠标器
 - 101 键键盘
 - 串口及并口
 - MIDI I/O 口
 - 游戏接口
2. 软件平台
- 有多媒体扩展软件包的 Microsoft Windows
 - MS—DOS 3.1 以上

上述配置的 MPC 功能较弱, 不足以用来开发或播放较好的多媒体应用系统。

二、第二个层次的 MPC 技术规范标准

如上述, 这一标准是最有实用价值的。

1. 硬件

- CPU 最低要求为 25 MHz 的 486SX (或兼容机)
- RAM 4MB (建议 8MB)
- 外存 3.5 英寸 (1.44MB) 软驱, 160MB 以上硬盘驱动器

• CD—ROM 驱动器传输速率 300KB/秒, 在 150KB/秒下占的带宽不超过 CPU 带宽的 40%, 在传输速率 300K/秒下不超过 60%; 平均搜索时间小于或等于 400 毫秒; 平均无故障时间 10000 小时。

• 音频 带 CD—DA (CD 数字音频, 红皮书标准) 输出和音量控制的 CD—ROM 驱动器; 具有线性 PCM 采样的 16 位数模转换器 (DAC); 在缓冲区分空时, 有用中断进行 DMA 或 FIFO 缓冲传输的能力; 有 44.1, 22.05 和 11.025kHz 的采样频率, 立体声通道; 在 22.05 和 11.025kHz 下所占的带宽不超过 CPU 带宽的 10%, 在 44.1kHz 下不超过 15%; 具有线性 PCM 采样的 16 位模数转换器 (ADC), 有 44.1, 22.05 和 11.025kHz 的采样频率, 在缓冲区分满时, 有用中断进行 DMA 或 FIFO 缓冲传输的能力, 麦克风输入; 多声道内部声音合成器能力, 铃鼓能力, 同时发生的 6 个美妙的音调加上 2 个同时发生的强力的敲击音调; 内部混合三个 (推荐 4 个) 输入声源并在背板上输出线电平立体声的能力; 这 4 个声源为: CD (红皮书), 声音合成器, DAC (波形声音), 辅助声源 (推荐, 但不要求); 每个输入必须至少有 3

位音量控制（8个等级），对数式地逐渐减小；（强烈推荐用4位以上的音量控制）；如果所有声源为-10dB（消费者线电平：1毫瓦，600欧 = 0dB）无衰减，混合器将不切头，且输出在0—3dB之间；强烈推荐使用单独的声源，主数字音量控制寄存器和额外的线电平声源；推荐CD-ROM XA 声频能力；推荐支持采用ADPCM的IMA。

- 视频 分辨率为640×480，65536（64K）的彩色监视器；应能支持每秒15帧，256个颜色视频的多媒体应用。

- 用户输入要求 标准101键，IBM风格的键盘，DIN插头，或提供使用了键合并有相同功能的键盘；具有总线或串行插头的两按键鼠标器，至少有一个附加的通讯口是空着可用的。

- I/O 要求 标准的9针或25针串行口，可编程最高到9600波特，可切换的中断通道；具有中断能力的标准25针的双向并行口；带有入、出、转接端的一个MIDI口，对输入和FIFO必须有中断支持；IBM式的模拟或数字游戏棒接口。

2. 软件

MPC系统软件必须提供与Windows 3.0 plus Multimedia Extensions 和 Windows 3.1 二进制兼容性。

三、最小的全系统MPC配置

MPC第二个层次系统要求下述的组成部分，它们都必须满足前一节所列出的功能指标。

- CPU：25MHz 486SX 或兼容机。
- RAM：4MB（推荐8MB）。
- 外存：软驱，硬驱（最少160MB）。
- 光驱：倍速CD-ROM，带CD-DA，XA可用，可实现multisession。
- 音频：16位的DAC，16位的ADC，音乐合成器，板上的模拟音频混合。
- 视频：显示分辨率至少为640×480，具有65536种颜色。
- 输入：101键盘（或功能等同），两按键鼠标。
- I/O：串口、并口、MIDI I/O口、游戏棒接口。
- 系统软件：与Windows 3.0 plus Multimedia Extensions 或 Windows 3.1 二进制兼容。

四、最小升级套件配置

MPC第二个层次标准的最小升级套件配置如下：

- 光驱：倍速CD-ROM，带CD-DA，XA可用，可实现multisession。
- 音频：16位的DAC，16位的ADC，音乐合成器，板上的模拟音频混合。
- I/O：MIDI I/O口、游戏棒接口。
- 用升级套件提供系统软件——任选。

得到MPC的办法有两种。一种是购买成套的MPC，另一种办法是购买多媒体升级套件，使普通的PC机升级为MPC。实际上，因为MPC是一个标准，而不是计算机，故我们可以从不同的来源购买零部件，装配自己的MPC，只要符合该标准即可。升级套件典型的应包括CD-ROM驱动器和音频卡。至于如何升级的详情可参见后面的章节。

MPC的典型配置如图1—4所示。

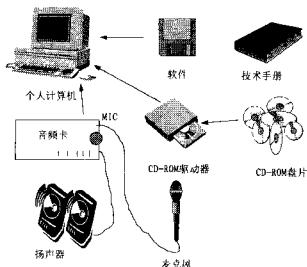


图 1 4 MPC 的典型配置

1—4 多媒体技术的发展简史

首先值得提出的是 1984 年 Apple 公司推出的 Macintosh 机引入了 Bitmap (位图, 或位映射) 的概念和用图符 (Icon) 来作为与用户的接口。在这个基础上进一步发展, 特别是在 1987 年 8 月引入了“超级卡” (Hypercard) 之后, 使 Macintosh 机成为使用方便、能处理多种信息媒体的计算机, 从而使它成为当时唯一可与 IBM PC 分庭抗礼的势力。

1986 年 3 月, Philips 和 Sony 联合推出了交互式紧凑光盘系统 CD—I (Compact Disc Interactive)。该系统把各种多媒体信息以数字化的形式存放在容量为 650 兆字节的只读光盘上, 用户可通过交互的方式来播放光盘中的内容。

1987 年 3 月, RCA 公司推出了交互式数字视频系统 DVI (Digital Video Interactive)。它以计算机技术为基础, 用标准光盘片来存储和检索静止图像, 活动图像, 声音和其它数据。RCA 后来把 DVI 技术卖给了通用电气公司, 后者又把这一技术卖给了 Intel 公司, 1989 年 3 月, Intel 宣布将把 DVI 技术开发成一种可以普及的商品, 包括把他们研制的 DVI 芯片装在 IBM PS/2 上。DVI 系统曾在 91 年美国计算机大展上荣获“Comdex 91”最佳奖。

随着多媒体技术的发展, 为建立相应的标准, 1990 年 11 月由 Philips 等 14 家厂商组成的多媒体市场协会应运而生, 今后要用 MPC 这个标志, 就要遵守这个协会所定的技术规范。MPC 标准的第一个层次是在一台 10MHz 286AT 的基础上增加硬盘和 CD—ROM (不久, 这个标准修改为采用 16MHz 的 386SX)。

1991 年第六届国际多媒体和 CD—ROM 大会上宣布的扩展结构体系标准 CD—ROM/XA, 目的是弥补原有标准在音频方面的缺陷。

在 1993 年推出的第二层次的 MPC 标准包括全活动的视频图像, 并将音频信号数字化时的采样量化值编码提高到 16 位。

JPEG 成为 ISO/IEC (国际标准化组织/国际电工委员会) 的 10918 号标准。

1994 年 11 月 MPEG-1 成为国际标准; 1994 年 11 月 MPEG-2 作为国际标准获得通过。

1-5 多媒体计算机系统的层次结构

多媒体计算机系统的层次结构如图 1-5 所示。这个系统与其他系统的结构在原则上是相通的, 最底层为硬件, 是系统的基础, 只是考虑到其“多媒体”的特性, 在每一层的具体内容上有所不同。

多媒体实时压缩和解压缩层。由于视频和音频信号要占用很大空间, 在处理时要对它进行压缩和解压缩, 而且要求处理速度极快。为此, 通常采用以专用芯片为基础的电路卡。许多集成电路厂商都在竞相开发这类产品, 现在已经形成了若干压缩与解压缩的标准。

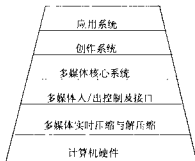


图 1-5 多媒体计算机系统的层次结构

多媒体入/出控制及接口层。与一般操作系统的功能类似, 它与多媒体硬件设备打交道, 驱动、控制这些硬设备, 并提供软件接口, 以便于高层软件调用。

多媒体核心系统 (Multimedia Kernel System) 层。它基本上就是多媒体操作系统。由于未来的计算机都要向多媒体方面发展, 谁能占领这一市场, 谁就能取得主动权。INTEL 和 IBM 公司为 DVI 系统开发的 AVSS 和 AVK 就起到了这样的作用。Microsoft 公司的 Windows, Apple 公司的 Quick Time 及 System 7.0 等都在争取占领这一市场。

创作系统 (Authoring System) 层, 有时也包含一些开发工具。这一层是为了方便开发者和用户开发应用系统用的。通常除了编辑功能外, 它还具有播放功能, 可用它控制多媒体外设。Microsoft 公司的 Multimedia Extension 1.0 和 Apple 公司的 Quick Time 就是这样的系统。

对于应用系统开发者来说, 一开始最好慎重选择、配置一套硬、软件都比较齐全的系统作为开发平台。但是, 在早期, 由于条件的限制, 并不一定能完全做到, 实际上在条件不成熟的情况下也可创造条件开展工作。作者在深圳蛇口新欣软件产业有限公司工作时, 1989 年开始, 从香港科学博物馆、宇航博物馆、台湾某自然科学博物馆等处承接了一大批多媒体应用项目订单, 当时, 除了使用 Macintosh 机的开发条件较好外, 大量基于 PC 机的项目, 就是在只有一些音、视频电路卡、触摸屏和简单的支持软件的基础上进行开发的。作为自主开发的结果, 除了开发出了数十个大大小小的多媒体应用项目外, 还创建了一套自己的多媒体软

件开发平台，做了许多有价值的工作，并取得了许多有益的经验。当然，时过境迁，现在已完全没有必要这样做了。

1—6 我国多媒体的发展现状

我国多媒体技术和应用的发展起于八十年代末，大致分为以下几个阶段：

1. 1989 年开始，比较多的工作集中在多媒体应用系统的开发上，从国外引进了一些类似于后来被称作声霸卡和视霸卡的声频卡和视频卡，在计算机上开发多媒体的应用系统。

2. 稍后，为了提高开发应用系统的效率和质量，人们开始注意创建自己的开发平台、著作工具、编辑软件等；有的更进一步引入国外的器件和部分技术，开发声频卡和视频卡之类的硬件产品。除了开发者自用之外，其中的某些佳品还被进一步产品化后作为正规产品销售。

3. 从 1992 年初开始，我国的多媒体热快速升温。那时，人们除了看到种种应用实例而受到启发之外，也开始可以从市场上（主要在深圳与北京）买到支持多媒体应用的板卡级产品，如声霸卡和视霸卡之类。人们发现，板卡厂商所提供的驱动程序和库函数之类支撑软件已比较解决问题，通过 C 语言之类的程序设计语言编程调用，已经很容易实现一些简单的多媒体应用系统。

4. 1993 年以后，随着应用水平的提高，特别是由于板卡销售利润丰厚所带来的竞争，使板卡的价格直线下降，由于成本降低，多媒体的应用进一步地得到推广；同时，多媒体技术水平有较大的提高（如关键的压缩和解压缩技术，平台技术，多媒体数据库技术等）；国内的有关的产品，如开发平台、多媒体数据库、支撑工具、音视频板卡、触摸屏等也以不同的规模推向市场。

5. 1994 年下半年开始，MPEG 及 JPEG 技术和有关产品的推广；CD—ROM、V—CD 及其播放器、播放卡的推广；多媒体计算机在市场大受青睐，特别是开始以前所未有的速度进入家庭；点播电视系统的开发；信息高速公路的发展和多媒体通讯在国内受到极大的重视；多媒体技术的研究正往深层发展……。这些都标志着我国多媒体事业的发展正在逐步加速，并上升到一个新的、可喜的阶段。

第二章 多媒体信息的计算机表示

多种媒体要进入计算机，其核心问题是数字化，在计算机中各种多媒体信息均变成二进制数字来进行处理。在讨论这一问题时，并非“一视同仁”，而是有所侧重。这一章对今后处理多媒体信息来说是一非常重要的基础部分，虽然许多内容都属于概念性的东西，但只有在概念上对这些组成多媒体的“元素”有了比较清楚的了解，今后才能高效率、高质量地开展相应的工作。

2—1 文 字

在计算机中，文字(Text，视场合有时也称之为“正文”)用二进制的编码表示，即使使用不同的二进制编码来代表不同的文字。

2—1—1 西文

用ASCII码表示。ASCII是美国信息交换标准代码(American Standard Code for Information Exchange)的简称。它是一个由7个二进制位组成的字符编码系统，包括大、小写字母，标点符号，阿拉伯数字、数学符号、控制字符等共128个字符。在全世界，ASCII码得到了最广泛的应用。

2—1—2 汉字

与西文不同，汉字的数目很多，我们常用的汉字就有几千个。现常用的二级汉字就用了两个字节，而表示更多的汉字时，还要用更多的字节。在我国使用计算机，汉字是必须考虑的重要因素，我国还为此制订了若干国家标准。例如：

GB2312-80 标准(1980年制订，用 2×7 个二进制位表示一个汉字，共有6763个汉字和850个符号)

ISO10646 标准(16位编码，32位、2万多汉字、中、日、韩均可、已经被批准为国际标准)

2—2 音 频

通常，音频(Audio)所指的是大约在15—20000Hz的频率范围，而实际上“音频”常常被作为“音频信号”或“声音”的同义语。在计算机中，处理音频信号前都要首先对其进行数字化。我们可以把各种声源(如麦克风、磁带录音、无线电和电视广播、CD等)所产生的声音数字化。数字化的声音是一种“采样的声音”。对声音进行数字化时，首先是采样，然后是量化。

2-2-1 音频信号的数字化处理

采样(Sampling, 有时也称之为“取样”)的作用是把时间上连续的信号, 变成在时间上不连续的信号序列(即通常的 A/D 变换, 将模拟信号转换成数字信号), 如图 2-1 所示。为了方便, 我们采用的是最简单的正弦波形, 这对说明概念来说已经够了, 但实际的声波波形要复杂得多。根据采样定理, 采样的频率只要高于信号最高频率的 2 倍就行了(为概念上的简单计, 这里只考虑低通型的信号)。采样的频率越高, 声音“回放”出来的质量也越高, 但是要求的存储容量也越大。

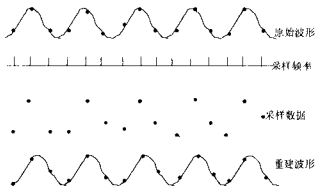


图 2-1 音频信号的数字化处理

另一个问题是“量化”, 即把采样所得的值(通常为反映某一瞬间声波幅度的电压值)加以数字化, 即用二进制来表示。显然, 用来表示一个电压值的二进制位数越多, 分辨率和质量越高。国际标准的语音编码采用 8 位(即有 256 个量化级)。

在多媒体中, 对于 CD 级的音频质量最常用的有三种采样频率: 44.1kHz, 22.05kHz, 11.025kHz。而量化(即分辨率)的位数采用 8 位和 16 位, 后者对应 65536 个量化级。

总的说来, 要求声音的质量越高, 为了保存这一段声音的相应的文件就越大, 也就是要求的存储空间越大。表 2-1 示出了采样频率、分辨率与所要求的文件大小的对应关系。

对单声道(MONO), 决定数字录音文件大小的公式为:

$$S = R \times D \times (r/8) \times 1$$

其中, S 表示文件大小(Size), 以字节计

R—采样速率(Sampling Rate), 以 kHz 计, 也可叫采样频率

D—录音的时间(Duration of Recording), 以秒计

r—分辨率(resolution), 以二进制位计, 如 8 位, 16 位

公式中的数字 1 表示对应的单声道。公式中的“除 8”是为了把二进制位换算成以字节作为单位。

例如, 在采样速率为 22.05kHz, 分辨率为 8 位, 单声道, 录音的时间长度为 10 秒的情况下, 文件的大小 S 为

$$S = 22050 \times 10 \times (8/8) \times 1 \\ = 220.5 \text{ KB}$$

对立体声，决定数字录音文件大小的公式与单声道的情况类似（仍以字节为单位）：

$$S = R \times D \times (r/8) \times 2$$

其中各符号的含义与上式相同，唯一不同的是乘以数字 2，它对应立体声，也就是说，立体声的文件大小为单声道的两倍。

例如，在采样速率为 44.1kHz，分辨率为 16 位，立体声（上述条件符合 CD 质量的红皮书音频标准，消费级级的音频压缩盘即按此录制），录音的时间长度为 10 秒的情况下，文件的大小 S 为

$$S = 44100 \times 10 \times (16/8) \times 2 \\ = 1764 \text{ KB}$$

从上可知，音频的数字化需要占用很大的空间，故解决音频信号的压缩问题是十分必要的。

表 2-1 采样速率、分辨率与存储空间的关系

采样速率	分辨率	立体声或单声道	1 分钟所需字节
44.1kHz	16 位	立体声	10.5MB
44.1kHz	16 位	单声道	5.25MB
44.1kHz	8 位	立体声	5.25MB
44.1kHz	8 位	单声道	2.6MB
22.05kHz	16 位	立体声	5.25MB
22.05kHz	16 位	单声道	2.5MB
22.05kHz	8 位	立体声	2.6MB
22.05kHz	8 位	单声道	1.3MB

获取的声音即各种自然声源产生的声音。对它们，可以采用前面所述的办法来使之数字化的。常用的一些标准为：

PCM —— 脉码调制 (CCITT G.711 —— 8K 采样，8 位量化精度，64KB/S)

DPCM —— 压缩脉冲编码

ADPCM —— 自适应压缩脉冲编码 (CCITT G.721)

2-2-2 MIDI

声音有两类：一种是获取的声音，一种是合成声音。合成声音可以是音乐或语言，合成声音与 MIDI 有紧密的联系，已形成标准，而合成语言则还未形成标准。

MIDI (Musical Instrument Digital Interface —— 乐器数字接口) 是 80 年代初提出来的，用作电子乐器和计算机之间的通信标准。MIDI 信息实际上是一段音乐的描述，当 MIDI 信息通过一个音乐或声音合成器 (synthesizer) 进行播放时，该合成器对一系列的 MIDI 信息进行解释，然后产生出相应的一段音乐或声音。MIDI 提供详细描述乐谱的协议（如音符、音调、使用什么乐器等）。

必须注意的是，MIDI 并不是数字化的声音，它仅仅是以数字形式存储的音乐的一种速记表示。MIDI 文件是用来记录音乐“动作”（如按下钢琴键，抬起踏板等）的一串与时间有关

的命令。简明的 MIDI 信息可以引起产生复杂的声音或在乐器或声音合成器上产生出美妙的音乐。因此, MIDI 文件比等效的(按每秒发出的声音计)数字化波形文件小得多。图 2-2 示出了 MIDI 的连接方法。在计算机上作曲很简单,只要有计算机、声音合成器和作曲软件,即可通过键盘逐一键入各种音符、音色等。如果不满意,可不断进行修改,或重新再来,直至满意,并作为一个音乐文件存入硬盘中为止。以后,通过播放软件就可以对这个音乐文件进行播放。

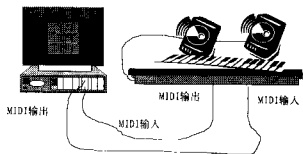


图 2-2 MIDI 的连接方法

2-2-3 数字化声音和 MIDI 的比较

与 MIDI 数据相比,数字化的声音是声音的实际表示。它代表了声音的瞬时幅度。因为它与设备无关,每次播放时它都发出相同的声音。从这一点看,它的一致性较好,但代价较高,因其数据文件要求较大的存储空间。

MIDI 数据是与设备有关的,即 MIDI 音乐文件所产生的声音与用来回放的特定的 MIDI 设备有关。

与数字化声音相比, MIDI 有几个优点,和两个大的缺点。其优点为:

- 文件紧凑,所占空间小, MIDI 文件的大小与回放质量完全无关。通常, MIDI 文件比 CD 质量的数字化声音文件小 200 到 1000 倍,它不占许多内、外存空间和 CPU 资源。

- 在某些情况下,如果所用的 MIDI 声源较好, MIDI 有可能发出比数字化声音更好的质量。

- 在不需要改变音调或降低音质的情况下,可以改变 MIDI 文件的长度(通过改变其速度)。MIDI 数据是完全可编辑的,我们可以用多种方法来处理它的每一个细节(可达亚毫秒的精度),而在处理数字化声音时,这些方法却完全用不上。

其缺点为:

- 因 MIDI 数据并不是声音,仅当 MIDI 回放设备与产生时所指定设备相同时,回放的结果才是精确的。

- MIDI 不能很容易地用来回放语言对话。

总的说来,数字化声音最重要的优点是其重放质量的一致性,可靠性比较好,可自始至终保证质量,而 MIDI 在这一点上则比较差。因此,在多媒体中,数字化声音用得比 MIDI 更多。

此外，还有两个重要的理由使人们对数字化声音感兴趣：

- 在应用软件和系统支持方面都有更多的选择，不管对 Macintosh 还是 Windows 平台均如此。

- 为了创建数字化声音所要求的准备与编程工作不需要掌握许多音乐理论知识，而 MIDI 则要求比较多。

2—2—4 在数字化声音和 MIDI 之间如何进行选择

1. 通常，下列情况适宜于用 MIDI：

- 由于没有足够的 RAM、硬盘空间或 CPU 处理能力使数字化声音不能工作。
- 具有高质量的声源。
- 对回放的硬件有完全的控制。
- 没有语言对话的需要。

2. 通常，下列情况适宜于用数字化声音：

- 对回放硬件没有完全的控制。
- 有足够的计算资源处理数字文件。
- 有语言对话的需要。

2—3 图 像

图像 (Image) 是多媒体中的可视元素。静态图像 (Still Image) 可大可小，也可占满整个荧光屏。静态图像在计算机中可以由两种方法产生：一种是位图 (bitmap)，一种是矢量画 (vector—drawn)。位图有时也称之为“位映射”。

2—3—1 位图

位图适用于逼真照片或要求精细细节的图像。而矢量画则适用于直线、方框、圆圈、多边形，以及其他可以用角度、座标和距离来进行数学表示的那些图形。两种类型的图像以不同的文件格式存储，它们之间可以互相转换，也可以从一种应用转换到另外一种应用，或者从一种计算机平台转换到另外一种计算机平台。通常，图像文件总是以压缩的方式进行存储的，以节省内存和磁盘空间。一般说来，静态图像是多媒体项目中最重要的元素。

在计算机的荧光屏或其他显示和打印设备上，分辨率的最小元素是独立的“点”(dot)，这些点也常常被称之为像素 (pixel，也称之为象素、象元等)。每个像素可以具有不同的颜色和亮度，象素也是能独立地赋予颜色和亮度的最小单位。

一个图像由若干个点 (或象素) 组成。通常，内存中划出一部分用作显示存储器 (也称帧存储器)，其中存放了与屏幕画面上的每一个象素——对应的一个个矩阵 (读者不难从这里体会到“位映射”的含义，这对其内部如何工作的机理很有帮助)；亦即，有多少个象素就有多少个矩阵。矩阵中的每一个元素就是象素值，象素值反映了对应象素的某些特性，而这个矩阵就称之为“位图”。简而言之：位图是一个用来描述象素的简单的信息矩阵。如果说得较“复杂”一点，也可以说：位图是一个用来描述构成一个图像的所有象素的特征的信息矩阵。

单色的（仅有黑、白两种颜色）可用一维矩阵（即一位的位图）来表示，而更多的颜色则要用较大的“深度”（depth），即用多位信息来表示，称之为“颜色深度”（要注意的是，这种“颜色深度”与我们平常所说的颜色深浅无关）。例如，4 位可以表示 16 种颜色，8 位可以表示 256 种颜色，16 位可以表示 32768 种颜色，而 24 位则可以表示 1600 多万种颜色（达到了“照片逼真”的水平），如此等等。图 2—2 用简单的模型来加以说明。图中，有 $4 \times 4 = 16$ 个象素，亦即这个图像仅有 16 个象素（实际上要多得多），但这种概念模型有助于读者理解。用 1 位位图时每个象素可以有黑白两种颜色，而用 24 位位图时每个象素则可以有 1600 多万种颜色。

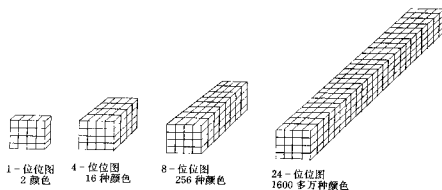


图 2—3 位图构成图像的概念模型

有三种产生位图的方法：

- 用画图程序得出。
- 用荧光屏抓取程序从荧光屏上直接抓取，然后把它加到画图程序或应用程序中。
- 用扫描仪或数字化的视频图像抓取设备从照片、艺术作品或电视图像抓取。

另外一种方法是购买现成的图像库。这种办法在国外十分流行。厂商把各种图像数字化以后放在磁盘或光盘中，像普通软件一样销售。由于是专业化的开发，规模化的生产，所以质量好而价格也比较适中。

用来处理位图的软件既有简单的，也有复杂的。随机提供的、免费的软件如：Macintosh 机的 HyperCard, PC/Windows 的 Paintbrush, 较为复杂的如：Aldus 公司的 SuperPaint, Computer Association 公司的 Cricket Paint 等等。这些软件可以处理上百，上千万种颜色，但实际上可以处理颜色的多少还取决于所用的视频卡和监视器等硬件。

现在我们来讨论一下图像的抓取和编辑问题。我们在荧光屏上看到的图像实际上是存储在视频存储器（它是内存的一部分）中的数字位图，大约每 1/60 秒或更短的时间内（取决于所用监视器的扫描速度）更新一次。抓取图像的最简单的方法是在观察到适当图像的某一瞬间按下键盘上的某一个键，它引起视频位图在格式上转换到我们将要使用的另外一种位图（在格式上不同）。在 Macintosh 和 Windows 环境中都有一个叫做 Clipboard 的内存区，当你在应用系统中进行剪裁和复制时，正文、图像这样的数据都暂时存储在这个内存区里。在 Windows 中，当我们按下“PRINT SCREEN”时，荧光屏上图像的一个副本就进到 Clipboard 中，

然后,我们可以从 Clipboard 中将抓取的位图“粘贴”到像 Paintbrush (Windows 中的一个绘图实用程序)这样的应用中,或者,我们可以使用主程序组中的 Clipboard 实用程序,将它作为一个 CLP 文件存起来。

在处理位图时,图像编辑程序可使我们得到许多创造性的功能。例如把两幅照片奇妙而天衣无缝地合成在一起。

此外,图像编辑工具可以使图像改变和变形。例如,一张彩色照片上的一朵红玫瑰可以变成一朵紫玫瑰,如果喜欢,也可以把它变成自然界中难以或见不到的兰色或黑色的玫瑰,也可以改变它的形状。

变形 (Morphing) 可以用来管理静态图像,或者创造一种有趣的,十分不寻常的,有生气的图形变形,从而产生另外一种特殊的效果。它允许人们把两个图像平滑地溶合在一起,看起来就好像是一个图像溶入了另一个图像中一样。例如,通过计算机一定算法的处理,它可以把我们熟知的一个人物,不知不觉地,逐渐地变成了另一个我们熟悉的人物。这样,常常可以产生某种逗人发笑的效果。

2-3-2 矢量画

大多数多媒体著作系统都提供了若干矢量画对象供用,这些矢量画对象有直线、矩形、椭圆、多边形以及正文等。

矢量画用得很多。比如,计算机辅助设计 (CAD) 系统中常用矢量画对象系统来创造一些十分复杂的几何图形;三维动画也使用矢量画对象。

如何作矢量画呢? 矢量是用其两个端点位置来描述的一条直线。例如,一个矩形可以简单地定义如下:

```
RECT 0, 0, 200, 200
```

其含意是,软件将画出一个矩形,它先从荧光屏的左上角开始,水平走 200 个像素,再往下走 200 个像素,最后画成一个正方形。下面这个描述:

```
RECT 0, 0, 200, 200, RED, BLUE
```

将画出一个同样的正方形,只不过其边框线是红色,方框中填以兰色。软件为什么要这样解释? 这完全是在设计软件时所制订的语法所规定的。

2-3-3 矢量画与位图的比较

在上述对一个彩色正方形的描述中,所用的数字和字母仅用了不到 30 个字节。如果经过压缩,所用的字节数还要更少。另一方面,若用未经压缩的位图,对于同样的一个正方形,对黑白图像而言 (每个像素仅有 1 位颜色深度),将占用 5000 个字节 ($200 \times 200 / 8$),其中,数字 8 表示一个字节有 8 位;对 256 色的图像而言 (每个像素有 8 位颜色深度),则将要求 40K 字节 ($(200 \times 200 / 8) \times 8$)。

从性能的角度上看,当在荧光屏上画了许多对象时,由于每个对象的大小、位置和其他特性都需要计算,等待荧光屏的刷新就需要很长的时间,速度变慢。例如,由 500 个单独的直线和矩形构成的图像就比仅由处理几个对象组成的图像花的时间多很多。这就是说,对于复杂的图像,用位图可以比用矢量画对象来作图得到更快的荧光屏刷新速度。

2-3-4 矢量画图像与位图图像之间的转换

大多数的绘图程序提供几种文件格式来保存图像。如果需要，可以在保存图画时，把由矢量画构成的图画变换成位图。

把位图变换成绘图对象则比较困难。但是，也有许多程序和实用程序，它们被用来计算图像中位图的边界，或者彩色块的形状，然后得出用来描述该图像的多边形对象。这一过程叫做“自动跟踪”，在某些集成了位图和矢量画对象的创作系统中（如 SuperCard）就提供这种功能。

2-3-5 图像与图形的简单比较

除了图像之外，我们还常常提到图形（Graphics）。对人而言，图形和图像是没有区别的。图形是用许多基本图形（如上述的直线、矩形、椭圆、多边形等）来描述的，而图像是用像素来描述的。二者各有其优缺点，用图形表示时用的数据量较小，处理容易，但用图形的办法来表示包含有许多细节的照片之类却有困难。在计算机中，大多数的图是用图像表示的。

有若干图形标准，下面列出几个有代表性的

- CGM (Computer Graphics Metafile) —— 计算机图形元语言
- STEP —— 工程图纸在计算机中的描述
- SPDL —— 标准页面描述语言标准

2-3-6 关于颜色

颜色是多媒体的至关重要的组成部分。这里主要讨论在计算机监视器荧光屏上显示的颜色是如何产生的。

因为人的眼睛对红、绿、蓝光敏感，通过调节这三种原色的组合成分使人的眼睛和大脑感受到各种颜色。这种颜色是心理上的，而不是物理上的颜色。比如，我们在计算机荧光屏上感受到的橙色，实际上是红光和绿光两种频率的复合，而不是当我们在阳光下看到实际的水果橙子时，频谱中真实的频率。所有这些因素使计算机的颜色处理起来十分错综复杂。

红绿蓝通常称为 RGB（即 Red、Green、Blue），它们的不同组合使我们可以感受到各种颜色。表 2-2 示出了这一情况。

表 2-2 RGB 组合与感知的颜色之间的关系

RGB 组合	感知的颜色
仅有红色	红色
仅有绿色	绿色
仅有蓝色	蓝色
红与绿(减去蓝)	黄色
红与蓝(减去绿)	紫色
绿与蓝(减去红)	青色
红、绿、蓝	白色
红、绿、蓝均无	黑色