



教育部人才培养模式改革和开放教育试点教材

计算机应用专业系列教材

多媒体技术基础实验

陈明 袁薇 编



37-33
6

中央广播电视大学出版社

前 言

中央广播电视大学计算机应用专业应用软件开发方向开设有多媒体技术基础、Visual BASIC 编辑与应用、软件工程、软件工具与环境 4 门必修方向课程。

为了提高学生学习质量，每门课程都配置相应的实验教材。《多媒体技术基础实验》包含 5 个实验：

1. 多媒体信息的表示
2. 多媒体信息的压缩
3. 多媒体开发环境与工具
4. 多媒体硬件安装与调试
5. 多媒体项目开发实践

考虑到这是一个大专层次的计算机教材，所设置的实验都比较简单实用。学生通过实验，能够进一步加深对理论的理解，提高实际动手能力。

本书在编写过程中得到中央电大各级领导和中央电大出版社领导及编辑人员的关心、指导和大力支持，也得到很多同仁的帮助，在此表示衷心感谢。

由于时间紧迫，能力有限，不足之处请批评指正。

陈 明

2000. 6. 30 于北京

实验一 多媒体信息的表示

声音媒体是多媒体计算机系统中的重要媒体信息。无论是用内置于计算机中的喇叭发声，还是使用声卡连接到音箱发声，数字化声音不仅使计算机能演奏悦耳动听的音乐、制造某种气氛和效果，也是各种多媒体软件、网络电话广泛使用的常用媒体。本实验通过使用声卡及其控制软件来录制、编辑、播放声音文件，让学生体会音频信号转换为数字化信号的处理过程。

一、实验目的

通过本实验，掌握音频和图像在计算机内的表示原理与方法。

所需知识点：音频数字化
位图与矢量图
图像文件

二、实验内容

1. 通过麦克风用 Windows 的录音机录制一段音频信号(可以是语音，或是音乐)，并储存于硬盘中。录制时分别用 44.1kHz, 22.05kHz, 11.025kHz 的采样频率对声波进行采样，每个采样点的量化位数分别选用 8 位、16 位，录制 10s（即秒，后同）的立体声节目。先计算其波形文件所需的存储容量，并与录音机录制的效果进行比较。

2. 计算出 64 位位图所能表示的颜色数目。

3. 计算出 1cm²黑白矩形所占用的存储空间（分别计算位图与矢量图的情况并比较计算结果）。

三、实验要求

硬件基本配置：Intel Pentium 120 CPU 或以上级别处理器；
大于 16Mb 的内存；
8 位以上的 DirectX 兼容声卡。

软件要求：Windows 录音机（Windows 95/98 操作系统）；
声卡随带的软件。

1. 实验学时 2 小时。
2. 熟读实验预备知识，掌握音频、位图和矢量图的概念。

3. 学会用 Windows 录音机来录音。
4. 事先完成所有计算题。
5. 写出实验报告。

四、实验预备知识

1. 音频数字化

音频 (Audio) 的频率范围大约在 20~20 000Hz, 也称为“音频信号”或“声音”, 属于听觉类媒体, 主要分为波形声音、语音和音乐。

音频具有音调、音强、音色三要素。音调与频率有关, 音强与幅度有关, 音色由混入基音的泛音所决定。

音频称为连续型时基媒体。因为声音是连续的, 它的产生离不开时间, 声音数据具有很强的前后相关性, 其数据量大、实时性强。

(1) 数字音频

数字音频区别于模拟音频, 在声音的录制和播放等方面有很大不同。计算机内的音频必须是数字形式的, 要实现音频数字化。即把模拟音频的波形信号转换成有限个数字表示的离散序列。

数字音频优于模拟音频的特点是信号在多次转录后不衰减, 保真度好, 动态范围大。在这一处理技术中, 要考虑采样、量化和编码的问题。

把模拟音频信号转换成计算机中的数字音频的过程如图 1-1 所示:

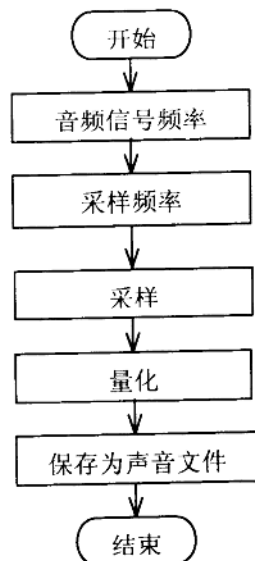


图 1-1 音频信号处理过程流程

① 影响数字音频质量的参数：采样频率、量化级、声道个数

采样的功能是将模拟信号转换成数字信号。其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 模数转换。采样频率是对声音波形每秒钟进行采样的次数。采样频率越高，声音“回放”出来的质量也越高，但是要求的存储容量也越大。

最常用的三种采样频率是：44.1kHz，22.05kHz，11.025kHz，其中前两项是最常采用的频率。

量化级也称量化数据位数，是每个采样点能表示的数据范围，其二进制位常用的有 8 位、16 位和 32 位。以 8 位的量化级为例，每个采样点可以表示 2^8 ，256 个不同的量化值；量化级为 16 位，则对应应有 2^{16} ，65 536 个不同的量化值。量化级越高，则数据量越大，音质越好。

声道通常分为单声道、立体声（双声道）、四声道。单声道一次只能生成一个声波数据。立体声每次生成两个声波数据，并在录制过程中分配到两个独立的声道输出，产生空间感的声音效果。四声道使声音从四个方向发出，加强了声音的环绕效果。

② 声音文件

声音文件是计算机中存储声音信息的文件。一般说来，要求声音的质量越高，则采样频率、量化级、声道个数也越高，存储这一段声音的磁盘空间也就越大，即声音文件越大。主教材第二章表 2-1 给出了采样频率、量化级与相应的声音文件大小的对应关系。

对于单声道，声音文件大小的公式为：

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

其中， S (Size) 表示文件大小，单位为 byte；

R 表示采样速率 (Sampling Rate)，单位为 kHz，也可叫采样频率；

D 表示录音的时间 (Duration of Recording)，单位为 s；

r 表示分辨率 (resolution)，即量化级。单位为二进制位 (bit)，如 8 位、16 位等；

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，一个字节等于八位二进制位。

对于立体声，将上式中的 1 换成 2，即立体声文件大小为单声道文件的两倍。

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

例如，如果采样速率为 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，消费者级的音频压缩盘即按此录制，录音的时间长度为 10s 的情况下，文件的大小 S 为： $S=(44100 \times 10 \times 16/8) \times 2 \text{kb} \times 1764 \text{ kb}$

可见，对音频数字化信号进行压缩是十分必要的。

在多媒体技术中，存储声音信息的文件格式主要有：WAV 文件、VOC 文件、MIDI 文件、AIF 文件、SNO 文件、RMI 文件、MP3 文件等。其中 MP3 文件是当前最热门的文件格式，它是 MPEG Layer 3 的缩写，它的压缩比较大，可达 1:12。

(2) 乐器数字接口 (MIDI)

MIDI 是 Musical Instrument Digital Interface (乐器数字接口) 的英文缩写，是世界上主要电子乐器制造厂商建立起来的一个数字音乐的通信标准。

MIDI 不仅提供了详细描述乐谱的协议（音符、音调、使用什么乐器等），还规定了各厂

家生产的电子乐器和计算机之间连接的电缆和硬件接口标准及设备间数据传输的规程。任何电子乐器，只要有处理 MIDI 信息的处理器并配以合适的硬件接口，均可成为一个 MIDI 设备。简明的 MIDI 信息可以产生复杂的声音或在乐器、声音合成器上产生出美妙的音乐。

MIDI 文件是一系列数字化的乐谱指令，由音符序列、定时及合成音色的乐器定义组成。所以 MIDI 文件比数字化的波形文件要小得多。

例如，一个典型的 8 位、22.05kHz 的波形文件，记录 1.8s 的声音需要 316.8k 空间，而一个 2 分钟的 MIDI 文件仅需 8k 的空间。由于 MIDI 文件比波形文件的长度小、安装方便，所以在设计多媒体应用和播放指定音乐时具有很大的灵活性。

与数字化声音文件比较，MIDI 文件有下列特点：

① 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 文件大小是 CD 质量的数字化声音文件的 1/1 000~1/200，它不占用较多的内存、外存空间和 CPU 资源。

② 设备相关性。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。而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。

③ 在不需要改变音调或降低音质的情况下，可以通过改变其速度来改变 MIDI 文件的长度。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，而在处理数字化声音时，这些方法却完全用不上。

④ 制作 MIDI 文件不仅需要掌握许多音乐理论知识，相关的设备准备与编程工作也比创建数字化声音文件要多。

2. 位图与矢量图

(1) 位图图像

位图图像是指在空间和亮度上已经离散化的图像。一幅位图图像可以考虑为一个矩阵，矩阵中的任一元素对应图像中的一个点，相应的值表示该点的灰度或颜色等级。矩阵的元素为像素，每个像素可以具有不同的颜色和亮度，像素也是能独立地赋予颜色和亮度的最小单位。

位图图像适用于逼真照片或要求精细细节的图像。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，位图图像是多媒体项目中最重要元素。

(2) 位图的概念

一个位图图像由若干个像素组成。通常，内存中划出一部分空间用作显示存储器，也称帧存储器，其中存放了与屏幕画面上的每一个像素一一对应的一个个矩阵。矩阵中的每一个元素就是像素值，像素值反映了对应像素的某些特性，而这个矩阵就称之为位图。

简而言之，位图是一个用来描述像素的简单的信息矩阵。如果说单色的（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多位信息来表示。例如，4 位可以表示 2^4 即 16 种颜色，8 位可以表示 256 种颜色，16 位可以表示 32 768 种颜色，而 24 位则可以表示 1 600 多万种颜色（可达到“照片逼真”的水平），如此等等。

图像的数据量很大，需经过压缩后再进行存储和传输。因此，研究压缩算法是非常重要的。

要的。压缩分有无损压缩、有损压缩，后者对图像的质量有影响，通常采用一种折衷的方案。如果原始采集的图像质量不好或者由于外界噪声影响而产生杂色、杂斑等，那么就应该采用图像优化技术。通过对图像的增强、噪声过滤、畸变校正、亮度调整和色度调整等，可获得满意的图像。

(3) 矢量图形

图形是一种抽象化的图像，是对图像按某个标准进行分析而产生的结果。它不直接描述数据的每一点，而是描述产生这些点的过程和方法。通常，将图形称之为矢量图形。

矢量图形是用一个指令集合来描述的。这些指令用来描述构成一幅图的直线、矩形、圆、圆弧、曲线等的形状、位置、颜色等各种属性和参数。显示时，需要相应的软件读取和解释这些指令，并将其转变为屏幕上所显示的形状和颜色。由于大多数情况下不用对图像上的每一点进行量化保存，因此需要的存储量较小。

产生矢量图形的程序通常称为绘图程序，它可以分别产生和操作矢量图形及各个片段，并可任意移动、缩小、放大、旋转和扭曲各个部分，即使相互覆盖或重叠，也依然保持各自的特性。矢量图形主要用于线形的图画、美术字、工程制图等。

(4) 矢量图与位图的比较

空间：矢量图形需要的存储量较小，而位图图像需要的存储量较大。

性能：对于复杂的图像，用位图比用矢量图速度快。

3. 图像文件的格式

Windows 使用设备无关位图 DIB (Device Independent Bitmaps) 作为其通用的图像文件格式。DIB 可以是独立的，也可以隐藏在一个叫做“资源交换文件格式” RIFF (Resource Interchange File Format) 的文件中。在 Windows 中，RIFF 实际上是所有多媒体开发中大家比较喜欢使用的文件类型，因为这一文件包含了许多种文件类型，如位图、MIDI 乐谱及格式化的正文等。

Windows 中最常用的图像文件格式是 DIB, BMP, PCX 和 TIFF。这些格式之间，可以用实用程序 CONVERT 来进行相互转换。

五、实验步骤——音频数字化

1. 检测所有设备的完整性。

声卡在计算机中的主要作用是处理声音文件、控制音调、提供 MIDI 接口等。声音的采集（即录制）、编辑加工（如加入背景音乐、删除某段声音等）、声音的回放（即播放）都离不开声卡。

音源，如收音机、家用录音机、CD 机等。又如用于录制音源的**麦克风**，播放声音文件的**音箱**。

将上述设备与声卡正确连接，完成硬件准备工作。

2. 插好电源，启动计算机。调试使声卡正常工作的软件。

在 Windows 98 中，单击(开始) → (设置) → (控制面板)，双击(多媒体)图标，打开“多媒体属性”对话框，选择“音频”选项卡。如图 1-2 所示，在“回放”、“录音”框中选择输出、

输入的首选设备。

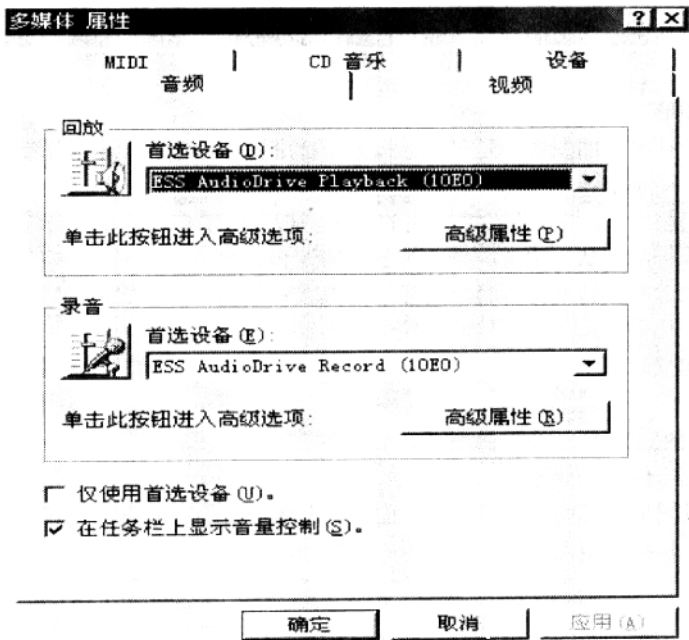


图 1 2

调整音量

双击位于任务栏右下角的喇叭，打开“音量控制”对话框，如图 1-3 所示。调整音量，注意(全部静音)的前面没有打“√”，否则的话，话筒和麦克风将没有声音。

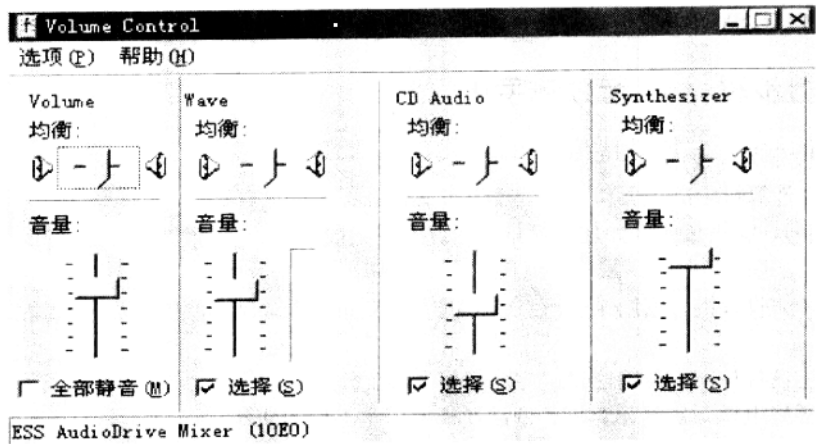


图 1-3

3. 单击(开始) → (程序) → (附件) → (娱乐) → (录音机)，启动 Windows 的录音机。



图 1-4

Windows 录音机录制音频文件时，一次能录制的时间为 60s。在录制过程中，所用的时间屏幕上会有显示，录制时间大于 60s 后，按(录音)键可继续录制。录音结束时，按(停止)键。

对于 Windows 95 操作系统，操作为：

4. 双击“控制面板”中(多媒体)图标，打开(多媒体属性)对话框，选择“音频”选项卡。

5. 在“录音”框中选择(自定义)按钮，可以更改录音的收音质量的“采样频率”、“量化位数”、“声道”、“每秒所需字节”。如图 1-5 所示。

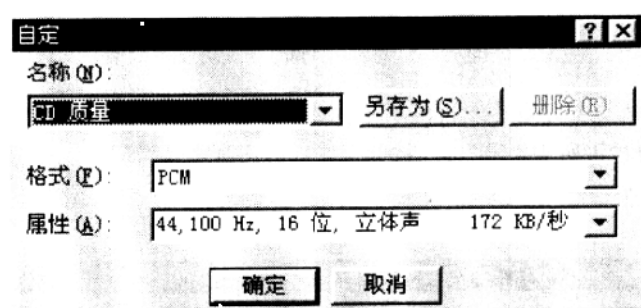


图 1-5

6. 按要求完成好声音文件的配置，按下(确定)按钮，设置结束。

7. 按“录音机”的(录音)按钮录音。录音结束后，将文件另存为 WAV 格式。

8. 重复步骤 4 至 7 直到做完实验为止，记录音频文件的存储容量。

对于 Windows 98 操作系统，操作为：

4. 单击(编辑) → (音频属性)，打开“音频属性”对话框。在“录音”框中选择(高级属性)按钮，打开“高级音频属性”对话框。调整“采样率转换质量”的尺度，按(确定)按钮。

5. 单击 [文件] → (新建)，建立一个新的声音文件。

6. 单击 [文件] → (属性)，打开“声音的属性”对话框，单击(开始转换)按钮，打开“选择声音”对话框。

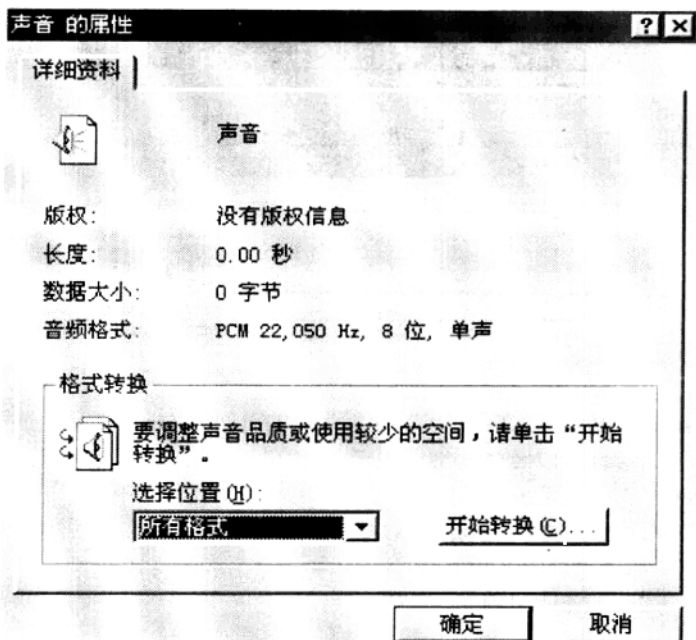


图 1-6

7. 在“属性”下拉列表中可以更改录音的收音质量的“采样频率”、“量化位数”、“声道”、“每秒钟所需字节”；按要求完成好声音文件的配置，按下[确定]按钮，设置结束。如图 1-7 所示。

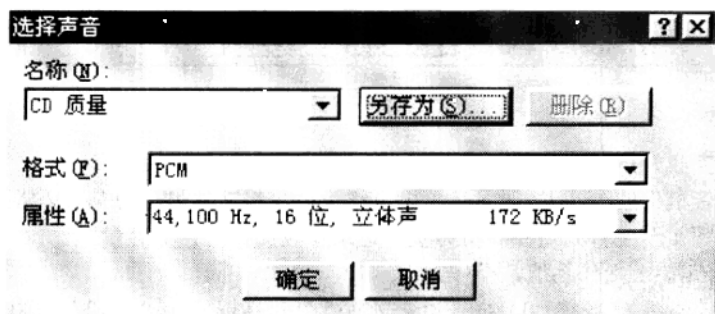


图 1-7

8. 按“录音机”的[录音]按钮录音。录音结束后，单击录音机菜单[文件]→[另存为]，打开“另存为”对话框，将文件另存为 WAV 格式。如图 1-8 所示。

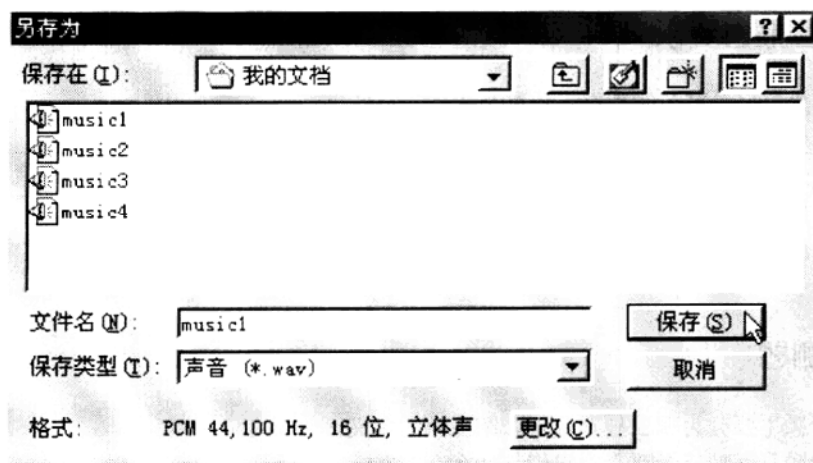


图 1-8

9. 重复步骤 5 至 8 直到做完实验为止，记录音频文件的存储容量。

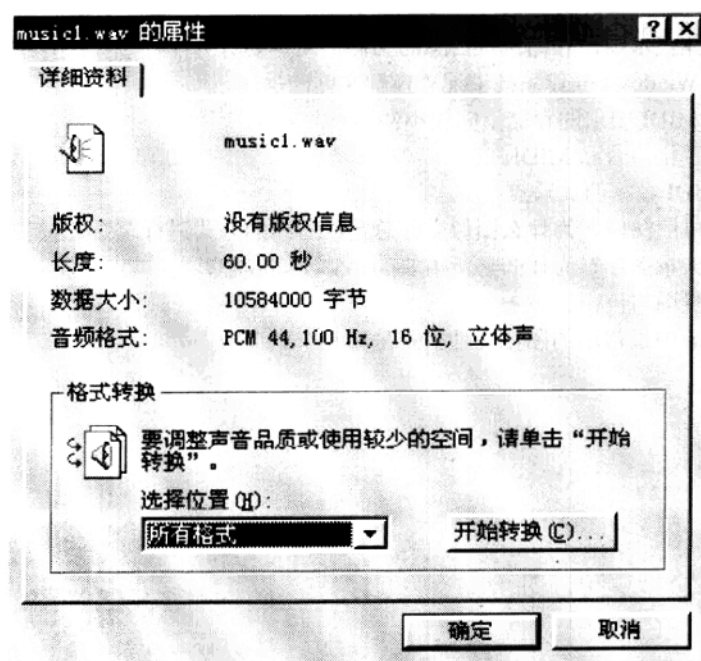


图 1-9 music1.wav 文件的详细资料

六、实验报告要求

1. 事先写好已准备的数据，计算公式，计算结果。
2. 比较计算结果与实验结果，写出结论。
3. 记录做实验过程中的体会。
4. 提出对该实验的意见和改进建议。
5. 提出迄今为止在多媒体领域还未出现的，很可能将要出现的想象中的新的一种媒体（可选）。

七、自测题

1. 音频主要分为哪三种？分别介绍各个的含义。
2. 音频具有哪三个要素？它们与什么有关？
3. 数字音频是什么？音频数字化怎样实现？在这一处理技术中，要考虑什么问题？
4. 描述一个音频信号转换成在计算机中的表示的过程。
5. 什么是采样？对采样频率的要求是什么？
6. 采样分辨率是什么？影响分辨率高低的因素是什么？
7. 写出决定单声道和双声道（立体）数字录音文件大小的公式。
8. 在多媒体技术中，存储声音信息的文件格式主要有哪几种？分别介绍其特点。
9. 描述在 Windows 操作系统中配置声音文件的操作。
10. 为什么 MIDI 要求的存储空间较小？
11. 比较数字化声音和 MIDI。
12. 指出 MIDI 数据的优缺点。
13. 什么是位图图像？为什么图像文件总是以压缩的方式进行存储？
14. 什么是矢量图形？为什么它所需要的存储空间较小？
15. 比较矢量图与位图。
16. Windows 中最常用的图像文件格式是哪几个？请分别介绍。

实验二 多媒体信息的压缩

一、实验目的

了解压缩技术对多媒体信息的传送带来的好处，并学习比较流行压缩工具的使用。

所需知识点：音频信号的压缩

视频图像的压缩

二、实验内容

使用现有比较流行的数据压缩工具（如 WinZip）来压缩某一段音频和视频信号并比较压缩文件的大小。

常用的文件压缩类软件有：

ARJ：ARJ 是目前使用非常广泛的一种压缩软件，它在 DOS 命令状态下执行，可以生成自解压文件，能够对根目录及其下面的所有目录及文件进行压缩。压缩速度也比较快。

RAR：RAR 弥补了 ARJ 的命令难以记忆、使用不方便的不足，是图形界面的压缩软件。它既可以管理档案文件，又可以管理一般文件，并具有高效的数据压缩及解压功能，具有 ARJ 的主要功能和非常高的数据压缩率。

PKZIP 和 PKUNZIP 是美国 PKWARE 公司生产的两个软件，在 DOS 状态下工作。PKZIP 的主要功能是对源文件进行压缩并建立压缩文件，它不仅具有完整的文件压缩功能，还可对它自己建立的压缩文件进行各种维护和提供压缩文件的技术信息。PKUNZIP 和 PKZIP 正好对应，是解压缩软件。

WinZip 是基于 Windows 的最著名的压缩软件。尤其是在 Internet 上奔波的玩家，几乎没有不知道 WinZip 的。这个软件拥有最多的用户，每当有新版本推出，就会吸引无数人下载。网址：<http://www.winzip.com>可以下载最新版本 WinZip 8.0 。

WinZip 的突出特点是：

- ① 灵活的向导和界面；
- ② 可以将文件压缩为自解压文件；
- ③ 支持 ARJ，ARC，ZIP，LZH 等压缩文件；
- ④ 方便的软件安装和反安装功能；
- ⑤ 提供密码保护。

WinArj：它并不是独立的压缩软件，仅仅是 DOS 下 ARJ 软件的 Windows 外壳，所有压缩和解压缩操作都需要 DOS 下的 ARJ 来完成。

ZipMagic: ZipMagic 98 是最新推出的一个压缩软件, 它的压缩功能比 WinZip 更加强大, 而且特色非常鲜明。它能够把压缩文件当成一个目录来管理, 这样杀毒工具就可以直接杀灭压缩文件中的病毒了。(有许多杀毒工具不能杀压缩文件里面的病毒)

三、实验要求

1. 实验学时 2 小时。
2. 熟读实验预备知识, 了解音频、视频信号的概念和各种压缩方法。
3. 提前学会几种流行的压缩工具的使用。
4. 做实验过程中, 注意 WinZip 对音频和视频信号的压缩率。
5. 写出实验报告。

四、实验预备知识

1. 音频信号的压缩编码

音频编码的目的在于压缩数据。只有当信源产生的信号具有冗余, 才能对其进行压缩。统计分析结果表明, 在语音信号中存在着多种冗余, 主要包括时域冗余和频域冗余。另外考虑到人的听觉机理特征, 也能对语音信号实行压缩。

在多媒体音频数据的存储和传输中, 数据压缩是必要的。通常数据压缩造成音频质量的下降、计算量的增加。因此在实施数据压缩时, 要在音频质量、数据量、计算复杂度三方面进行综合考虑。

在音频信号压缩编码之前, 需进行音频信息的数字化, 这部分内容请看实验一。

① 时域冗余度

A. 幅度的非均匀分布

统计表明, 语音中的小幅度样本比大幅度样本出现的概率要高。又由于通话中必然会有间隙, 因此出现了大量的低电平样本。此外, 实际讲话信号功率电平也趋向于出现在编码范围的较低电平端。因此, 语音信号取样值的幅度分布是非均匀的。

B. 样本间的相关

从语音波形的分析中可以看出, 在邻近样本之间取样数据存在最大的相关性。当取样频率为 8kHz 时, 相邻取样值间的相关系数大于 0.85, 甚至在相距 10 个样本之间, 还可有 0.3 左右的数量级。如果取样速率提高, 样本间的相关性将更强。因而根据这种较强的一维相关性, 利用 N 阶差分编码技术, 可以进行有效的数据压缩。

C. 周期之间的相关

语音信号虽与电视信号有许多相似之处, 但其最大的不同, 是语音信号的直流分量并不占主要成分。因为光信号是非负的, 而语音信号却可正可负。虽然语音信号需要一个电话通路提供整个 300Hz~3400Hz 的带宽, 但在特定的瞬间, 某一声音往往只是该频带内的少数频率成分在起作用。当声音中只存在少数几个频率时, 就会像某些振荡波形一样, 在周期与

周期之间,存在着一定的相关性,利用语音周期之间信息冗余度的编码器,比只利用邻近样本间的相关性的编码器效果要好,但要复杂得多。

D. 基音之间的相关

人的说话声音可分为两类:

第一类称为浊音(voiced sound),由声带振动产生,每一次振动使一股空气从肺部流进声道,激励声道的各股空气之间的间隔称为音调间隔或基音周期。一般而言,浊音产生于发元音及发某些辅音的后面部分。

第二类称为清音(unvoiced sound),一般又分成摩擦音和破裂音两种情况。前者用空气通过声道的狭窄部分而产生的湍流作为音源;后者声道在瞬间闭合,然后在气压被迫下迅速地放开而产生了破裂音源。语音从这些音源产生,传过声道再从口鼻送出。清音比浊音具有更大的随机性。

浊音波形不仅显示出上述的周期之间的冗余度,而且还展示了对应于音调间隔的长期重复波形,因此,对语音浊音部分编码的最有效的方法之一是对一个音调间隔波形来编码,并以其作为其他基音段的模板。男、女声的基音周期分别为 5ms~20ms 和 2.5ms~10ms,而典型的浊音约持续 100ms,一个单音中可能有 20~40 个音调周期。虽然音调周期间隔编码能大大降低编码率,但是检测基音有时却十分困难。如果对音调检测不准,便会产生奇怪的“非人音”。

E. 静止系数

两个人之间打电话,平均每人的讲话时间为通话总时间的一半,另一半时间听对方讲。听的时候一般不讲话,而即使是在讲话的时候,也会出现字、词、句之间的停顿。分析表明,话音间隙使得全双工话路的典型效率约为通话时间的 40% (或静止系数为 0.6)。显然,话音间隔本身就是一种冗余,若能正确检测出该静止段,便可“插空”传输更多的信息。

F. 长时自相关函数

上述样本、周期间的一些相关性,都是在 20ms 时间间隔内进行统计的所谓短时自相关。如果在较长的时间间隔(比如几十秒)进行统计,便得到长时自相关函数。长时统计表明,8kHz 的取样语音的相邻样本间,平均相关系数高达 0.9。

② 频域冗余度

A. 非均匀的长时功率谱密度

在相当长的时间间隔内进行统计平均,可得到长时功率谱密度函数,其功率谱呈现强烈的非平坦性。从统计的观点看,这意味着没有充分利用给定的频段,或者说有着固有的冗余度。特别地,功率谱的高频能量较低,这恰好对应于时域上相邻样本间的相关性。此外,再次可以看到,直流分量的能量并非最大。

B. 语音特有的短时功率谱密度

语音信号的短时功率谱,在某些频率上出现峰值,而在另一些频率上出现谷值。这些峰值频率,也就是能量较大的频率,通常称为共振峰频率。此频率不止一个,最主要的是第一和第二个,由它们决定了不同的语音特征。另外,整个谱也是随频率的增加而递减。更重要的是,整个功率谱的细节以基音频率为基础,形成了高次谐波结构。这都与电视信号类似,仅有的差异在于直流分量较小。

③ 人的听觉感知机理

A. 人的听觉具有掩蔽效应。当几个强弱不同的声音同时存在时,强声使弱声难以听见的现象称为同时掩蔽,它受掩蔽声音和被掩蔽声音之间的相对频率关系影响很大;声音在不同时间先后发生时,强声使其周围的弱声难以听见的现象称为异时掩蔽。

B. 人耳对不同频段的声音的敏感程度不同,通常对低频端比对高频端更敏感。即使是对同样声压级的声音,人耳实际感觉到的音量也是随频率而变化的。

C. 人耳对语音信号的相位变化不敏感。人耳听不到或感知极不灵敏的声音分量都不妨视为冗余。

④ 音频信息的数字化

声音信号是振幅随时间连续变化的模拟信号。由于计算机只能处理二进制编码的数字信号,因此在让计算机处理音频信息之前,必须先将模拟音频信号数字化。所谓音频信号的数字化,就是将模拟音频信号每隔一定时间间隔截取一段,并将所截取的信号振幅值用一组二进制脉冲序列表示,从而使连续的模拟音频信号等价地转换成离散的数字音频信号。

在数字化的过程中,以一定时间间隔截取模拟信号的处理称为采样。截取所得到的各段信号振幅值称为采样值,将采样值以二进制方式表示称为量化,而将这些二进制数组成脉冲序列的处理称为编码。音频信号的数字化不仅仅是计算机处理离散的信号,更重要的是,数字化音频信号具有极高的保真度。当前,传统的模拟音响系统和模拟视频系统正在逐渐被数字音响系统和数字视频系统取代就是最好的例证。

A. 数字化音频信息的质量与容量

采样,量化和编码技术是实现连续模拟音频信号离散数字化的关键技术。其中,采样频率、每个采样值的存储信息量及音频信息的声道数目是反映数字化音频信息质量和容量的三个重要因素。

采样频率,即单位时间内模拟音频信号被截取的段数,采样频率越高,截取的段数越多,数字化后的音频信息的质量就越高,但所需的信息存储量也越大。所以采样频率必须在声音质量和存储容量之间进行权衡。著名的香农(Shannon)采样定理指出:若对随时间连续变化的模拟信号波形用该信号所含的最高频率两倍的频率采样,则可基本保证原信号的质量。电话的传输频带大约为 300~3 400Hz,所以采用 7kHz 的采样频率便能基本保证语言信息的可理解度,男女语音的最高频率都不超过 10kHz,故用 20kHz 的采样频率便可真实地反映出各类语声的细微特征。而正常人耳听见的声音的频率上限为 20kHz,若用 40kHz 的采样频率,便可高保真地重现包括各类乐器在内的所有声音信息。多媒体计算机系统中采用的音频信息标准采样频率为 44.1kHz、22.05kHz 和 11.025kHz 三档,能满足各类不同应用的需要。

表示采样值的二进制数的位数称为量化位数。量化位数的多少反映了个采样值的精度。例如,4 位量化能反映采样值的 16 个等级,即最小能表示音频信号最大振幅值的 $1/16$;8 位量化能反映采样值的 256 个等级,其精度为音频信号最大振幅的 $1/256$;而 16 位量化的精度可为音频信号最大振幅的 $1/65\,536$ 。可见,量化位数越多,所得的量化值越接近于原始波形的采样值,精度就越高。但是,量化位数越多,所要求的信息存储量也就越大。

声音信息通道的个数是指声音记录为一组波形(单声道)还是两组波形(即双声道立体声),甚至更多组波形(多声道)。立体声听起来自然要比单声道的声音丰富、逼真得多,但需要成倍于单声道的信息存储容量。

从以上分析可以看出，数字化音频信息的质量与它的存储信息量具有正比关系。

B. 音频信号编码的分类

基于音频数据的统计特性、声学参数、人的听觉特性进行分类，音频信号的压缩方法有多种。如图 2-1 所示。

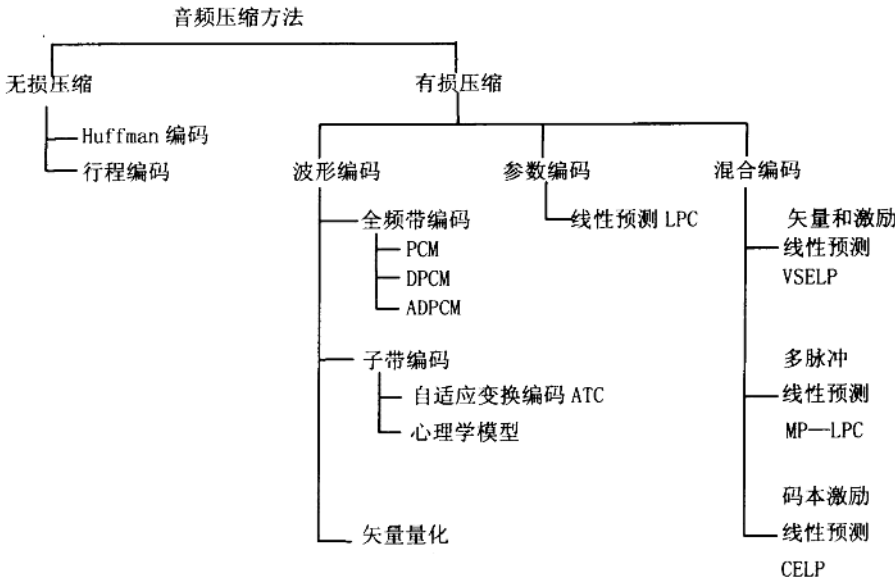


图 2-1 音频信号压缩方法

C. 音频信号压缩编码的基本原理

在多媒体计算机系统中，声音信号被编码成二进制数字序列，经传输和存储，最后由解码器将二进制编码恢复成原始的声音信号，如图 2-2 所示。

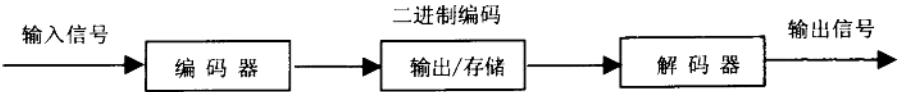


图 2-2 声音信号的处理流程

声音信号的编码方式大致可分为三大类，即波形编码方法、分析合成方法和混合编码方法。

D. 音频信号压缩编码技术

脉冲编码调制