



武汉体育学院试用教材

WU HAN TI YU XUE YUAN SHI YONG JIAO CAI

计算机音频制作

黄曦东 凌刚 周彤 编著



武汉体育学院教材委员会

计算机音频制作

黄曦东 凌 刚 周 彤 编著

武汉体育学院教材委员会

2005 年 6 月

前 言

《计算机音频制作》教材是根据我院专业建设和教材建设的总体目标,结合大学生素质教育及现代信息技术的飞速发展和广泛应用,面向 21 世纪,组织编写的。

本教材依据计算机音频制作的教学大纲,从声音、数字音频入手,详细介绍了 MIDI、音乐记谱软件 Finale、音乐器软件、数字音频处理软件的使用方法以及插件等音乐教育的相关软件。

本教材是针对体育院校体育信息技术系、体育艺术系本科生编写的,也可作为其他院系自由选修课程选用教材。

本教材的第一章、第二章、第五章、第六章由黄曦东撰写,第三章和第七章由凌刚撰写,第四章由黄曦东和周彤撰写。由于作者水平有限,加之时间仓促,难免有错误和不当之处,敬请专家和读者批评指正。

武汉体育学院教材委员会(计算机音频制作)教材小组

武汉体育学院体育信息技术系

2005 年 7 月

目 录

第一章 数字音频基础	1
1.1 声音	1
1.2 数字音频基础	3
1.2.1 数字电子的基本原理	3
1.2.2 数字音频基础	6
1.3 常见音频文件格式	8
1.4 软件分类	19
第二章 MIDI 基础	25
2.1 MIDI 基础知识教程	25
2.2 数字音乐合成原理	27
2.2.1 数字音乐合成原理	27
2.3 音源概述	34
2.3.1 软波表	34
2.3.2 XG100	36
2.3.3 RolandSC-88Pro	46
2.3.4 SoundFont	50
第三章 音乐记谱软件	58
3.1 Finale	58
第四章 音序器软件	92
4.1 CakewalksonarSONAR	92
第五章 数字音频处理软件	114
5.1 Cooleditpro	114
5.2 Acid	153
第六章 插件	176
6.1 插件	176
第七章 音乐教育及其他软件	191
7.1 Earmaster	191
7.2 Goldwave	191
7.3 超级解霸	200

第一章 数字音频基础

1.1 声音

我们都知道,声音是由于物体的振动而发生的,物体振动才能发声。

振动有快慢之分有的振动很快,有的振动较慢。振动的快慢我们用“周期”或“频率”来表示。振动着的物体,它的振动部分运动一个来回所需要的时间叫做周期,周期用“秒”作单位。如某物体的振动部分运动一个来回需要 2s 时间,则物体的振动周期 2s。周期越长,振动就越慢。在 1 秒钟内完成来回振动的次数,叫做频率,频率用“赫兹”作单位。1 秒钟振动 1 次,频率就是 1 赫兹 1 秒钟振动几次,频率就是几赫兹。有的振动的幅度大,有的振动的幅度小,振动的幅度是用“振幅”这个物理量来表示的,振幅就是振动物体偏离中心位置的最大距离。振幅越大,振动就越强。

声音必须靠物质来传播。

传声的媒质有气体,固体、液体。

传播的是一种运动,即声振动,而传声的媒质并没有随着声波前进。

声波在气体和液体中只有纵波。在固体中除了纵波以外,还可能有横波(质点振动的方向与声波传播的方向垂直),有时还有纵横波。

接收到声音还必须有声接收器,如拾音器(又称话筒)等,振动物体、介质、声接收器三者是接收到声音缺一不可的条件。

从本质上讲,声音是一种连续的波,称为声波。

声源体发生振动会引起四周空气振荡,那种振荡方式就是声波。声波借助空气向四面八方传播。在开阔空间的空气中那种传播方式像逐渐吹大的肥皂泡,是一种球形的阵面波。声音是指可听声波的特殊情形,例如对于人耳的可听声波,当那种阵面波达到人耳位置的时候,人的听觉器官会有相应的声音感觉。

除了空气、水、金属、木头等也都能够传递声波,它们都是声波的良好媒质。在真空状态中声波就不能传播了。

正弦波是最简单的波动形式。优质的音叉振动发出声音的时候产生的是正弦声波。

正弦声波属于纯音。任何复杂的声波都是多种正弦波叠加而成的复合波,它们是有别于纯音的复合音。正弦波是各种复杂声波的基本单元。

扬声器、各种乐器以及人和动物的发音器官等都是声源体。地震震中、闪电源、

雨滴、刮风、随风飘动的树叶、昆虫的翅膀等各种可以活动的物体都可能是声源体。它们引起的声波都比正弦波复杂,属于复合波。地震产生多种复杂的波动,其中包括声波,实际上那种声波本身是人耳听不着的,它的频率太低了(例如 1Hz)。

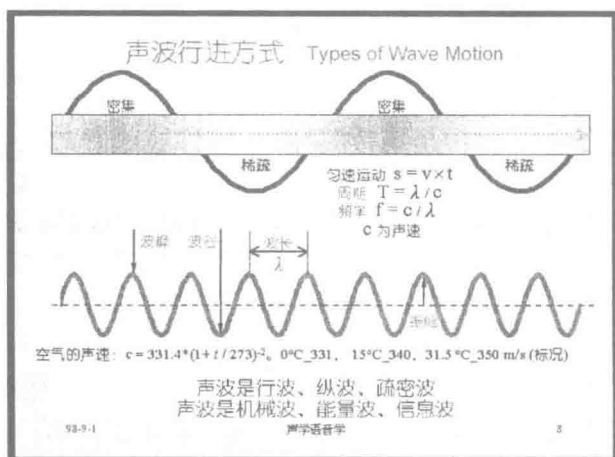


图 1-1

本图片(如图 1-1)的图解中只显示声波从左到右一个方向上传播的情形。我们可以设想,这里有一个很细很长的、内部是空气的直管子,现在正弦声波正在它里边行进。图中疏密交替指的是空气密度随声波改变的情形。背后那个正弦曲线并不表示声波上下运动;它是一种科学图解,表示空气密度的变化过程。也就是说那个正弦图解的纵轴表示密度的大小,横轴表示从声波前进的距离,曲线表示声波行进中空气密度的变化过程。那种变化过程呈正弦形。复合波里也有疏密交替,但是疏密交替的配合方式要复杂得多。

图解中的纵轴也可以理解为空气粒子进退的位移量,因此它一般都被称为振幅轴。我们经常拿水面波来类比声波。请注意,水面波是上下起伏的,那种位移的方向跟波前进的纵深方向不一致,所以称它为横波,它是横向振荡的。跟水面波不同的是,声波前进的时候,空气中相邻的粒子被压力变化驱动而前后振荡(位移),那种前后振荡的轨迹是跟声波前进的轨迹一致的,所以说声波是纵波,它是纵向波动的。

图解中的纵轴也可以理解为空气压力的变化。声波是大气压力之外的一种超压变化。

空气粒子振动的方式跟声源体振动的方式一致,当声波到达人的耳鼓的时候就引起耳鼓同样方式的振动。驱动耳鼓振动的能量来自声源体,它就是普通的机械能。不同的声音就是不同的振动方式,它们能够起区别不同信息的作用。人耳能够分辨风声、雨声和不同人的声音,也能分辨各种言语声,它们都是来自声源体的不同信息波。

请注意,声波不是冲击波,声波前进的过程是相邻空气粒子之间的接力赛,它们把

波动形式向前传递,它们自己仍旧在原地振荡,也就是说空气粒子并不跟着声波前进。

另外,即使没有其他声源体的作用,空气粒子总是在做无规则的震荡,或者说它们总是在骚动,它们激发起微弱的“白噪声”。绝对静寂的大气空间是不存在的。所谓背景噪声还包括自然界或人类生活环境里许多声源体杂乱的声音。

听觉角度来讲,声音主要由四种属性——高低、长短、强弱、色彩(即常说的音色)。其中:

音的高低是由发音源在一定时间内的振动次数(频率)来决定的。振动次数越多,频率越高,音也就越高。反之则低;

音的长短是由发音源振动时所持续的时间来决定的。持续的时间越长,音越长,反之则短;

音的强弱取决于由发音源的振动幅度(振幅),幅度越大,音越强,幅度小音就弱;而发音源的材质、形状及泛音数的多少则决定了音色。

声音的以上这四种属性,在音乐的表现中都是非常重要的,其中又以“高低”和“长短”最为重要。比如一首歌曲,本来是由人声演唱的,但是如果我们把它改编为器乐曲、并且减小音量,虽然改变了音的长短和音色,但是仍然可以很容易地分辨出它的旋律。但是,如果我们把这首乐曲中每个声音的长短、高低随意的改变,其原来的旋律马上会受到严重的破坏。

由于物体在振动的时候并不是总是规则的,所以声音又有“乐音”和“噪音”之分。规则振动发出的是乐音,反之则称为噪音。在音乐中,用的最多的是乐音,但是这并不代表噪音就不属于音乐的组成部分,乐队中的大部分节奏打击乐器所发出的声音就属于噪音,但它同样具有非常丰富的音乐表现力。

1.2 数字音频基础

1.2.1 数字电子的基本原理

比较仪:

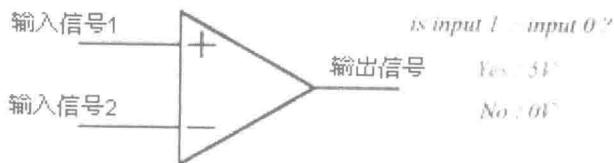


图 1-2 比较仪

会有两种情况发生:

如果输入信号 1 低于输入信号 2,那么输出信号电压为 0 伏。

如果输入信号 1 高于输入信号 2,那么输出信号电压为 5 伏。

那么,输出信号只可能是 0 伏或 5 伏,这就是数字信号的基本原理。可以把这个过程设想为比较仪提出一个问题:“输入信号 1 大于输入信号 2 吗?”那么 0V 则代表“非”或称“Boolean0”(Boolean 是二进制记数法中,内部表示为 0 或 1 的一种值);1V 则代表“是”或“Boolean1”。

模拟信号至数字信号的转换:

数字电子学的基本思路是把模拟信号编码组成 2 进制码。一套 2 进制码有许多的基本的“Bit”(比特)组成。例如:信号可以是 0 或者 1。

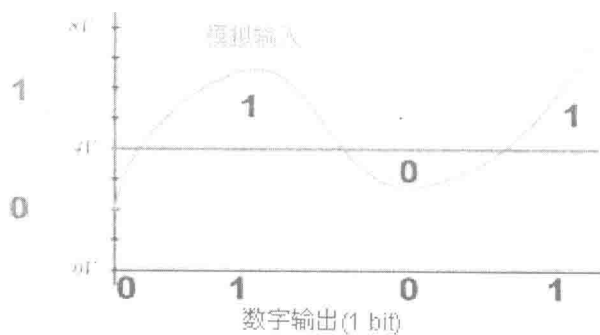


图 1-3

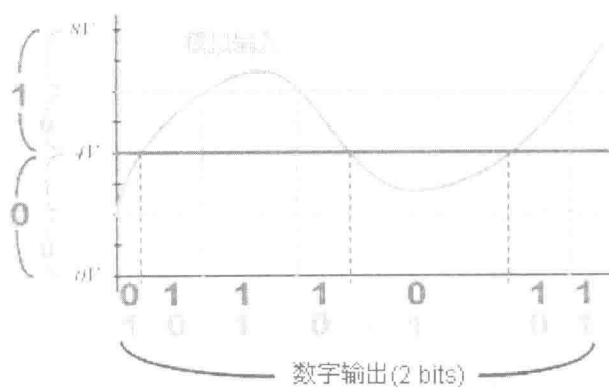


图 1-4

比较仪则提供了一个基本的模拟信号编码方式。我们可以设想在 0V 与 8V 之间的一些信号的变化。如果我们设定比较仪上的输入信号 2 为 4V,那么比较一就会告诉我们:输入信号可能会在 0—4V 之间或者 4—8V 之间。这个范围,我们可以把它划分为 2 个区:高区和低区。

观察上图(图 1-3),我们不难发现,用 1 比特来近似地“描述”出原信号,是非常

粗糙的。因为我们只取高区为 1、低区为 0。

经过改良,我们再加一些特殊的比较仪,这些比较仪能够判断出更详细的讯息,例如:多比特方案。以上图所示为 2 比特下的信号还原情况,以 4V 为界,信号变化可以被分成 4 个区域,无论高区还是低区都有 0 和 1。



图 1-5

依此看来,划分的区域越多,精确度则越高。也就是说比特率越高,精确度越高。即:

1bit:2 个模拟信号区

2bits:4 个

3bits:8

4bits:16

...

14bits:16384

15bits:32768

16bits:65536

可见,16bit 数字信号系统可以同时处理 65536 个区。这已经非常精确了:如果把 1 米长划分为 65536 格的话,那么每格长度仅为 0.015 毫米。人,已经很难分辨出比 16Bit 精度更细微的变化了。

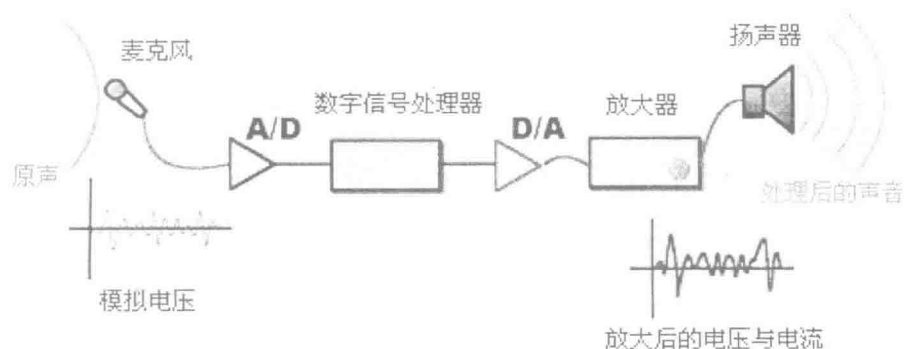


图 1-6

概括地说,数字电子的就是处理经编码的模拟信号。比特率越高,那么你得到的精确度则越高,对原信号的“描述”也就越逼真。2 进制算法是计算机的最基本算法。

因此,我们就可以用数字电子方式来“模仿”模拟电子的“行为”。

与模拟电子相比,数字信号最令人感兴趣是它没有模拟信号里的那些噪声。因为,它们本身仅仅是数字 0 和 1 而已。所以有人说:“数字化音频本身就是一种个性化的、复杂的艺术。”

1.2.2 数字音频基础

大家知道,无论现在的多媒体电脑功能如何强大,其内部也只能处理数字信息。而我们听到的声音都是模拟信号,怎样才能让电脑也能处理这些声音数据呢?还有,究竟模拟音频与数字音频有什么不同呢?数字音频究竟有些什么优点呢?这些都是我们下面所要介绍的。

把模拟音频转成数字音频,在电脑音乐里就称作采样,其过程所用到的主要硬件设备便是模拟/数字转换器(AnalogtoDigitalConverter,即 ADC)。采样的过程实际上是将通常的模拟音频信号的电信号转换成许多称作“比特(Bit)”的二进制码 0 和 1,这些 0 和 1 便构成了数字音频文件。如下图,图中的正弦曲线代表原始音频曲线;填了颜色的方格代表采样后得到的结果,二者越吻合说明采样结果越好。

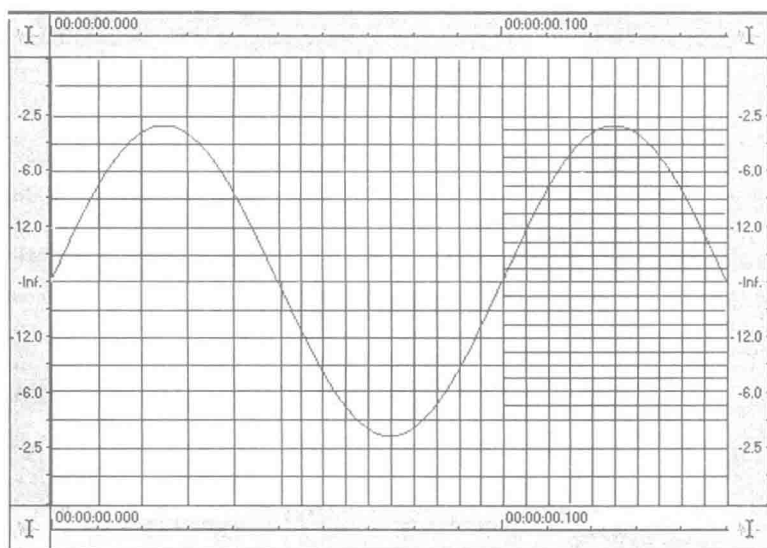


图 1-7

上图中的横坐标便是采样频率;纵坐标便是采样分辨率。图中的格子从左到右,逐渐加密,先是加大横坐标的密度,然后加大纵坐标的密度。显然,当横坐标的单位越小即两个采样时刻的间隔越小,则越有利于保持原始声音的真实情况,换句话说,采样的频率越大则音质越有保证;同理,当纵坐标的单位越小则越有利于音质的提高,即采样的位数越大越好。有一点请大家注意,8 位(8Bit)不是说把纵坐标分成 8 份,而是分成 $2^8=256$ 份;同理 16 位是把纵坐标分成 $2^{16}=65536$ 份;而 24 位则分

成 $2^{24}=16777216$ 份。现在我们来进行一个计算,看看一个数字音频文件的数据量到底有多大。假设我们是用 44.1kHz、16bit 来进行立体声(即两个声道)采样,即采样成标准的 CD 音质(也称作红皮书音频)。那么就是说,一秒钟内采样 44.1 千次,每次的数据量是 $16 \times 2 = 32\text{bit}$ (因为立体声是两个声道)。而大家知道,一个字节(Byte)含有 8 个位(Bit),那么一秒钟内的数据量便是 $44.1\text{k} \times 32\text{bit} / (8\text{bit}/\text{Byte}) = 176.4\text{kByte}$ 。一个汉字在电脑里占用两个字节,那么 176.4kB 的空间可以存储 $176.4\text{k}/2 = 88200$ 个汉字,也就是说一秒钟的数字音频数据量与近九万个汉字(一部中篇小说)的数据量相当。由此可见,数字音频文件的数据量是十分庞大的。

也许有人会问,为什么要把 CD 音质的采样频率规定成 44.1kHz 而不是其他的频率呢?这个问题问得好。44.1kHz 意味着每秒采样四万多下,这会不会太多了点呢?究竟每秒采样多少次才算合理呢?大家请看下图。图中,上半部分表示原始音频的波形;下半部分表示录制后的波形;红色的点表示采样点。

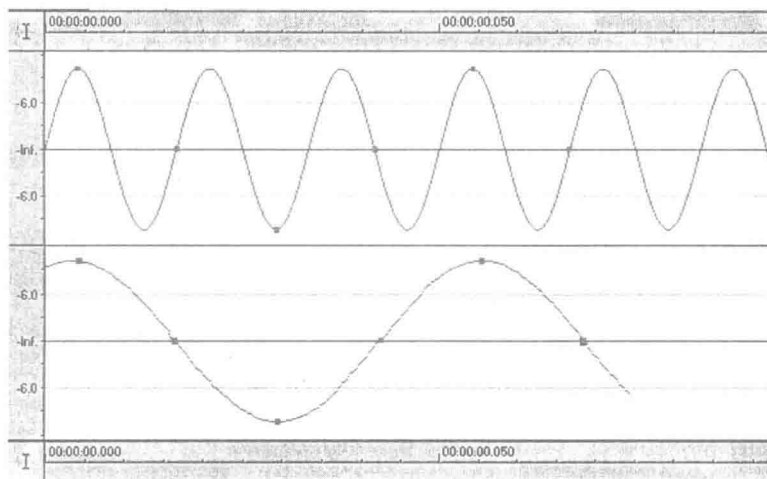


图 1-8

大家可以发现,上下波形之所以不吻合,是因为采样点不够多,或严谨一点说,是采样频率不够高。这种情况,我们称之为低频失真。

一个常见的低频失真的例子便是电影上车辆行驶时车轮转动的情况(一个典型的“马车轮”效应的例子)。你也许早已发现,飞快转动的车轮有时看起来似乎是静止不动甚至会向反方向转动(类似的情况也发生在直升飞机的翼片和螺旋桨上面)。关于合理的采样频率这一问题在 Nyquist(奈奎斯特)定理中早已有明确的答案:要想不产生低频失真,则采样频率至少得是录制的最高频率的两倍(上图中,采样频率只是录制频率的 $4/3$ 倍)。这个频率通常称作 Nyquist 极限。

在正常的音乐中,最高的音符也只不过 7kHz—8kHz,这似乎意味着 16kHz 的采样频率便已足够。其实这 7、8kHz 仅仅表示基音的音高,还有大量的泛音未包括在内,故

用这种方法来定采样频率是十分不科学的。其实,所谓“不失真”,换句话说便是“人们听不到失真”。人类的听力范围是 $20\text{Hz} - 20\text{kHz}$, 所以采样频率至少得是 $20\text{k} \times 2 = 40\text{kHz}$ 便可保证不产生低频失真。CD 音质的 44.1kHz 正是这样制定出来的(略高于 40kHz 是为了留有余地)。按照 Nyquist 定理,这样的采样频率可以保证即使是 22.05kHz 的超声波也不会产生低频失真。而音频的工业标准所规定的 48kHz 采样频率(如 DAT, DigitalAudioTape)则有更高的 Nyquist 极限,满足更苛刻的要求。

那么数字音频又是如何播放出来的呢? 首先,将这些由大量数字描述而成的音乐送到一个叫做数/模转换器(DigitaltoAnalogConverter,即 DAC)的线路里。它将数字回变成一系列相应的电压值,然后通过有助于稳定的保持线路,最后将信号由低通滤波器输出。这样,比较平缓的具有脉动电压的模拟信号可继续发送至放大器和扬声器,电流经过放大再转变成声音。

相对应的模拟音频又是怎样录制与播放的呢? 首先,声波通过麦克风,空气分子的振动转变为电信号的波动(数字录音也必需经过这一步)。录音磁头的电磁铁根据通过电流的大小而产生大小不同的磁场,磁场的变化情况会相应的记录在磁带上(实际上是磁带上的磁粉排列发生了变化),这样便完成录音过程。播放时,放音磁头读出印记在磁带上的磁场大小变化的情况(即磁粉的排列位置),并转变为相应的电信号。之后的情况与数字音频的播放完全类似,即这些波动的电信号(模拟信号)继续传送至放大器和扬声器,电信号重新转变为声音(即空气分子的振动)。

说到这里,我们可以理解数字录音的好处了。首先,录制好的音乐是以数字来储存的,而数字的传输错误率是相当低甚至是可以避免的,所以录制好的音乐可以多次复制而效果不减(这在制作过程中十分重要)。而模拟信号则每传输一次就失真一次。而且,模拟录音的本底噪音很大,要想满足严谨的录音要求则需要购买复杂而又昂贵的设备,操作也十分繁琐。况且,处理数字信息是电脑的拿手好戏,只需面对显示器,所有的工作都可以弹指一挥间完成。正是基于以上这些优点,使得建立一个家庭工作室(HomeStudio)显得非常具有吸引力,并且技术上也成为可能。

1.3 常见音频文件格式

目前的音频格式其实不少。每种格式都有自己的优缺点,在这样的情况下,究竟哪种音频格式适合自己的使用呢? 我们有必要做一个较全面的了解。先给大家介绍介绍常见的音频文件格式的特点。

CD 格式

把 CD 放进电脑,在播放软件的“打开文件类型”中,都可以看到 *.cda 格式,这就是 CD 音轨。标准 CD 格式也就是 44.1K 的采样频率,速率 $88\text{K}/\text{秒}$, 16 位量化位数,因为 CD 音轨可以说是近似无损的,因此它的声音基本上是忠于原声的。CD 光

盘可以在 CD 唱机中播放,也能用电脑里的各种播放软件来重放。一个 CD 音频文件是一个 *.cda 文件,这只是一个索引信息,并不是真正的包含声音信息,所以不论 CD 音乐的长短,在电脑上看到的“*.cda 文件”都是 44 字节长。注意:不能直接的复制 CD 格式的 *.cda 文件到硬盘上播放,需要使用抓音轨软件把 CD 格式的文件转换成 WAV,mp3,wma 等其他格式。

Wav——无损的音乐

是微软公司开发的一种声音文件格式,它符合 PIFFResourceInterchangeFile-Format 文件规范,用于保存 WINDOWS 平台的音频信息资源,被 WINDOWS 平台及其应用程序所支持,由于 Windows 本身的影响,这个格式已经成为了事实上的通用音频格式,它实际上是 Apple 电脑的 AIF 格式的克隆。

标准格式的 WAV 文件和 CD 格式一样,也是 44.1K 的采样频率,速率 88K/秒,16 位量化位数声音文件质量和 CD 一样,也是目前 PC 机上广为流行的声音文件格式,几乎所有的音频编辑软件都“认识”WAV 格式。

通常我们使用 WAV 格式都是用来保存一些没有压缩的音频,但实际上 WAV 格式的设计是非常灵活(非常复杂)的,该格式本身与任何媒体数据都不冲突,换句话说,只要有软件支持,你甚至可以在 WAV 格式里面存放图像。之所以能这样,是因为 WAV 文件里面存放的每一块数据都有自己独立的标识,通过这些标识可以告诉用户究竟这是什么数据。在 WINDOWS 平台上通过 ACM(AudioCompression-Manager)结构及相应的驱动程序(在这里通常称为 CODEC,编码/解码器),可以在 WAV 文件中存放超过 20 种的压缩格式,比如 ADPCM、GSM、CCITT G. 711、G. 723 等等,当然也包括 MP3 格式。

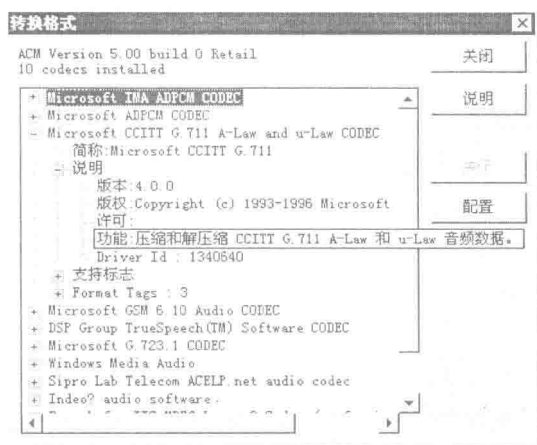


图 1-9

图表一使用音频编辑软件 WaveCN(<http://www.wavecn.com>)列出 Windows

安装了 ACM 驱动信息

MIDI 是乐器数字接口的英文缩写,是数字音乐/电子合成乐器国际标准。MIDI 目的是解决各种电子乐器间存在的兼容性问题。MIDI 规范不仅定义了电脑音乐程序,音乐合成器及其电子音乐设备交换音乐信号的方式,而且还规定了不同厂家的电子乐器与电脑连接的电缆和硬件及设备间数据传输的协议,可用于为不同乐器创建数字声音,能很容易的模拟钢琴,小提琴等传统乐器的声音。MIDI 本身并不能发出声音,它是一个协议,只包含用于产生特定声音的指令,而这些指令则包括调用何种 MIDI 设备的声音,声音的强弱及持续的时间等。电脑把这些指令交由声卡去合成相应的声音。最初,因为不同 MIDI 设备的乐器音色排列方法不一,所以会造成同一 MIDI 文件在不同的设备会出现完全不同的放声效果。为避免出现这种混乱情况,GM(GENERAL MIDI)标准被提出并得到了 WINDOW 操作系统的支持,得到了相当广泛得应用。它规定了前 128 中常用乐器的音色编排方式,例如 1 号是钢琴,66 号是萨克斯管等等,它实际上是对 midi 规范的补充。ROLAND 公司提出的 GB 标准兼容 GM 的基础上,对其进行了发展,增强了音乐的表现力——他提供比 GM 标准数量更多的打击乐器组合更多的特殊音响。GS 标准具有广泛的软硬件适应性,包括声卡,音乐爱好者的娱乐乐器啊到专业音乐器材等。后来,yamaha 公司又提出了基于 GM 标准的 XG 标准。相对于保存真实采样数据的声音文件,MIDI 文件显得更加紧凑,其文件的大小要比 WAV 文件小的多——一分钟的 WAV 文件约要占用 10MB 的硬盘空间,而已分钟的 MIDI 却只有区区的 3.4KB。现在,MIDI 已经成为电脑音乐的代名词。电脑播放 MIDI 文件时,有两种方法合成声音;FM 合成和波表合成。FM 合成是通过多个频率的声音混合来模拟乐器的声音,波表合成是将乐器的声音样本存储在声卡波形表中,播放时从波形表中取出来,产生声音。采用波表合成技术,可以产生更逼真的声音。MIDI 文件有几个变通的格式,其中 CMF 文件是随声卡一起使用的音乐文件,于 MIDI 文件非常相似,只是文件头略有差别;另一种 MIDI 文件是 WINDOWS 使用的 RIFF 文件的一种子格式,称为 RMID,扩展名为 RMI。

MP3 格式诞生于八十年代的德国,是 Fraunhofer—IIS 研究所(<http://www.iis.fhg.de/>)的研究成果。mp3 是第一个实用的有损音频压缩编码。

所谓的 MP3 也就是指的是 MPEG 标准中的音频部分,也就是 MPEG 音频层。MPEG 代表的是 MPEG 活动影音压缩标准,MPEG 音频文件指的是 MPEG 标准中的声音部分即 MPEG 音频层。MPEG 频文件根据压缩质量和编码复杂程度的不同可分为三层(MPEGAUDIOLAYER1/2/3 分别与 MP1。MP2 和 MP3 这三种声音文件相对应 MPEG 音频编码具有很高的压缩率,MP1 和 MP2 的压缩率分别为 4:1 和 6:1—8:1,而 MP3 的压缩率则高达 10:1—12:1,也就是说一分钟 CD 音质的音乐未经压缩需要 10MB 存储空间,而经过 MP3 压缩编码后只有 1MB 左右。因此,目前 INTERNET 上的音乐格式以 MP3 最为常见。MP3 为降低声音失真采取了名

为“感官编码技术”的编码算法:编码时先对音频文件进行频谱分析,基本保持低音频部分不失真,但是牺牲了声音文件中 12KHz 到 16KHz 高频部分的质量来换取文件的尺寸,然后用过滤器率掉噪音电平,接着通过量化的方式将剩下的每一位打散排列,最后形成具有较高压缩比的 MP3 文件,并使压缩后的文件在回放时能够达到比较接近原音源的声音效果。虽然他是一种有损压缩,但是它的最大优势是以极小的声音失真换来了较高的压缩比。

在 mp3 出现之前,一般的音频编码即使以有损方式进行压缩能达到 4 : 1 的压缩比例已经非常不错了。但是,mp3 可以实现 12 : 1 的压缩比例,这使得 mp3 迅速地流行起来。mp3 之所以能够达到如此高的压缩比例同时又能保持相当不错的音质是因为利用了知觉音频编码技术,也就是利用了人耳的特性,基本保持低音频部分不失真,削减音乐中人耳听不到的成分,牺牲了声音文件中 12KHz 到 16KHz 高频部分的质量来换取文件的尺寸,同时尝试尽可能地维持原来的声音质量。

图表一 人耳等响度曲线图,mp3 的技术基础。

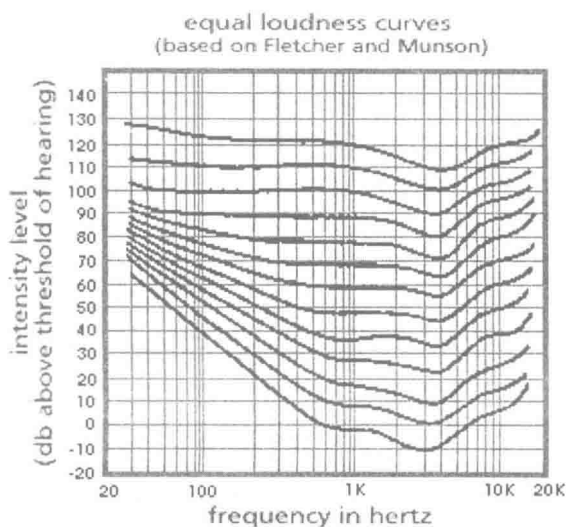


图 1-10

衡量 mp3 文件的压缩比例通常使用比特率来表示。这个术语的英文是 bps: bitpersecond,表示每 1 秒钟的音频可以用多少个二进制比特来表示。通常比特率越高,压缩文件就越大,但音乐中获得保留的成分就越多,音质就越好。由于比特率与文件大小音质的关系,所以后来又出现了 vbr(Variable Bitrate 可变比特率)方式编码的 mp3,这种编码方式的特点是可以根据编码的内容动态地选择合适的比特率,因此编码的结果是在保证了音质的同时又照顾了文件的大小,结果大受欢迎。其实 mp3 的编码标准本来就支持这种压缩方式,但是第一个将此功能实现的反而是一个第三方工具:曾经非常有名的 XingTechnology 公司(<http://www.xingtech.com>),当

年很多电脑上看 VCD 听 mp3 就靠这个公司的 XingPlayer)推出的 XingEncoder。

MP3 格式压缩音乐的采样频率有很多种,可以用 64Kbps 或更低的采样频率节省空间,也可以用 320Kbps 的标准达到极高的音质。我们用装有 FraunhoferL-ISMpegLyaer3 的 MP3 编码器(现在效果最好的编码器)MusicMatchJukebox6.0 在 128Kbps 的频率下编码一首 3 分钟的歌曲,得到 2.82MB 的 MP3 文件。采用缺省的 CBR(固定采样频率)技术可以以固定的频率采样一首歌曲,而 VBR(可变采样频率)则可以在音乐“忙”的时候加大采样的频率获取更高的音质,不过产生的 MP3 文件可能在某些播放器上无法播放。我们把 VBR 的级别设定成为与前面的 CBR 文件的音质基本一样,生成的 VBRMP3 文件为 2.9MB。

MP4:

MP3 问世不久,就凭这较高的压缩比 12:1 和较好的音质创造了一个全新的音乐领域,然而 MP[3 的开放性却最终不可避免的导致了版权之争,在这样的背景之下,文件更小,音质更佳,同时还能有效保护版权的 MP4 就应运而生了。MP3 和 MP4 之间其实并没有必然的联系,首先 MP3 是一种音频压缩的国际技术标准,而 MP4 确实一个商标的名称,其次,他采用的音频压缩技术也迥然不同,MP4 采用的是美国电话电报公司所研发的,以“知觉编码”为关键技术的 a2b 音乐压缩技术 <http://www.a2bmusic.com> 课讲压缩比成功的提高到 15:1,最大可达到 20:1 而不影响音乐的实际听感,同时 mp4 在加密和授权方面也做了特别设计,它有如下特点:

每首 mp4 乐曲就是一个扩展名为 .exe 的可执行文件。在 windows 里直接双击就可以运行播放十分方便。mp4 这个有点同时又是她的先天缺陷——容易感染电脑病毒!

更小的体积!更好的音质?想对先进的 a2b 音频压缩技术的采用,使 mp4 文件的大小仅为 mp3 的四分之三左右,从这个角度来看 mp4 更适合在 inter 上传播,而且据说音质也更胜一筹,但我怎么也没听出他比 mp3 的音质更为优越。

独特的数字水印。mp4 采用了名为“SOLANA”技术的数字水印。可方便的追踪和发现盗版发行行为。而且,任何针对 mp4 的非法解压行为都可能导致 mp4 原文件的损毁。

支持版权保护。mp4 乐曲还内置了包括与作品版权持有者相关的文字、图像等版权说明,即可说明版权。又表示了对作者和演唱者的尊重。

比较完善的功能。mp4 可独立调节左右声道音量控制,内置波形/分频动态音频显示和音乐管理器可支持多种彩色图像,网站连接及无限制的滚动显示文本。

RealAudio 主要适用于在网络上的在线音乐欣赏,这种文件格式几乎成了网络流媒体的代名词。RA、RMA 这两个文件类型就是 RealMedia 里面向音频方面的。它是由 RealNetworks 公司发明的,特点是可以在非常低的带宽下(低达 28.8kbps)提供足够好的音质让用户能在线聆听。

网络流媒体的道理其实非常简单,简单说就是将原来连续不断的音频分割成一个一个带有顺序标记的小数据包,将这些数据包通过网络进行传递,在接收的时候再将这些数据包重新按顺序组织起来播放。如果网络质量太差,有些数据包收不到或者延缓了到达,它就跳过这些数据包不播放,以保证用户在聆听的内容是基本连续的。

现在 real 的的文件格式主要有这么几种:有 RA(RealAudio)、RM(RealMedia, RealAudioG2)、RMX(RealAudioSecured),还有更多。这些格式的特点是可以随网络带宽的不同而改变声音的质量,在保证大多数人听到流畅声音的前提下,令带宽较富裕的听众获得较好的音质。随着网络带宽的普遍改善,Real 公司推出用于网络广播的、达到 CD 音质的格式。

网络流媒体的道理其实非常简单,简单说就是将原来连续不断的音频分割成一个一个带有顺序标记的小数据包,将这些数据包通过网络进行传递,在接收的时候再将这些数据包重新按顺序组织起来播放。如果网络质量太差,有些数据包收不到或者延缓了到达,它就跳过这些数据包不播放,以保证用户在聆听的内容是基本连续的。

WMA(WindowsMediaAudio)格式是来自于微软,音质要强于 MP3 格式,更远胜于 RA 格式,它和日本 YAMAHA 公司开发的 VQF 格式一样,是以减少数据流量但保持音质的方法来达到比 MP3 压缩率更高的目的,WMA 的压缩率一般都可以达到 1:18 左右,WMA 的另一个优点是内容提供商可以通过 DRM(DigitalRights-Management)方案如 WindowsMediaRightsManager7 加入防拷贝保护。这种内置了版权保护技术可以限制播放时间和播放次数甚至于播放的机器等等,这对被盗版搅得焦头乱额的音乐公司来说可是一个福音,另外 WMA 还支持音频流(Stream)技术,适合在网络上在线播放,而 Windows 操作系统和 WindowsMediaPlayer 的无缝捆绑让你只要安装了 windows 操作系统就可以直接播放 WMA 音乐。同一格式,音质好的可与 CD 媲美,压缩率较高的可用于网络广播。

WindowsMedia(<http://www.microsoft.com/windows/windowsmedia>)也是一种网络流媒体技术,本质上跟 RealMedia 是相同的。但 RealMedia 是有限开放的技术,比如 rtsp(RealTimeStreamProtocol 实时流协议)这样的网络传输协议是提交到网络工作组 RFC 网络协议集的其中一个(编号 RFC2326),而 WindowsMedia 则没有公开任何技术细节,据称是为了更好地进行版权保护,因此要完全封闭,还创造出一种名为 mms(Multi-MediaStream 多媒体流)的传输协议。

开放源代码的 OggVorbis(<http://www.xiph.org> 或 <http://www.vorbis.com>),作为开放源码向专利技术下的第一封战书出现在世人的面前。在经历了一些变故之后,OggVorbis 终于在 2002 年 7 月释出了 1.0 版本。由于开放源码的东西向来都缺少枪手为其摇旗呐喊,所以国内并不是有很多的人了解 OggVorbis。OggVorbis 是一种音频压缩格式,类似于 MP3 等现有的通过有损压缩算法进行音频压缩的音乐格式。但有一点不同的是,OggVorbis 格式是完全免费、开放源码且没有专