

数字音频处理技术

声音是人们进行信息交流的最好工具,自从计算机能够处理声音后,声音就成了重要的多媒体元素,成为多媒体信息的重要载体。声音在物理学上,的三个基本特性是频率、振幅和波形,对应到人耳的主观感觉就是音调、响度和音色。声音的数字化包括采样、量化和编码三个基本过程,数字化声音质量的好坏取决于采样的频率和采样的精度。

在数字媒体中,数字音频是携带信息的重要媒体。音频信号可分为两类:语音信号和非语音信号。非语音信号又可分为音乐和杂音。非语音信号的特点是不具有复杂的语音和语法信息,信息量低,识别简单。语音是语言的载体。语言是人类社会特有的一种信息系统,是社会交流工具的符号。

随着数字媒体技术与内容的发展,数字音频处理技术受到人们的高度重视,数字音频不仅作为单一媒体形式,同时也与其他媒体(例如图文、视频等)构成多媒体形式,对提升和丰富数字媒体内容起着举足轻重的作用,并得到了广泛的应用。例如,影视作品的配音、配乐,游戏的音响效果,虚拟现实中的声音模拟,数字出版的有声输出,语音识别,声音操控等。

5.1 音频的概念及特性

声音是通过一定介质(如空气、水等)传播的一种连续振动的波,声音被看成一种波动的能量,也称为声波。声波是由物体振动所产生并在介质中传播的一种波,具有一定的能量。同时在物理学上,一般用声音的三个基本特性来描述声音,即频率、振幅和波形。生理学上,声音是指声波作用于听觉器官所引起的一种主观感觉。声音的主观感觉是听觉的主观属性,属于心理学范畴。人的感觉不像话筒的测试系统那样绝对化,人类对物理量的响应通常与所描述的物理单位量并不一致,因为这里存在一个心理物理量的问题。这就是为什么会出现人们对声音量的主观描述,如响度、音调、音色和音长等。

尽管这两个关于声音的理解含义有所不同,但它们之间有一定的内在联系。在物理学上,声音的三个基本特性(频率、振幅和波形),对应到人耳的主观感觉就是音调、响度和音色。

具体来说,所谓频率即发声物体在振动时单位时间内的振动次数,单位为赫兹(Hz)。一般来说,物体振动越快,频率就越高,人感受到的音调也越高,反之亦然。这也是为什么把声音称为音频的主要原因。

振幅是指发声物体在振动时偏离中心位置的幅度,代表发声物体振动时动能、势能的大小。振幅是由物体振动时所产生声音的能量或声波压力的大小所决定的。声能或声压越大,引起人耳主观感觉到的响度也越大。

音色是指声音的纯度,它由声波的波形形状所决定。即使两个声音的振幅和频率都一样,也就是说它们的音调高低、声音强弱都相同,但若它们的波形不一样,听起来也会有明显的区别。例如,听音乐时,能分辨出胡琴、小提琴和钢琴等乐器。日常生活中人们听到的多是复合音,单纯的纯音是很少的。实验室的音频发生器和耳科医生用来检查病人听觉用的音叉能发出纯音。

声音在生活中是无处不在的,不但能够为人们传递信息,而且能够带来感官的愉悦。因此,人们通常对声音的分类也有多种不同方法。

按照人耳可听到的频率范围,声音可分为超声、次声和正常声。人耳不是对所有物体的振动都能听得见。物体振动次数过低或过高,人耳都不能感受。人耳可感受声音频率的范围为20~20 000Hz。低于20Hz的信号为亚音信号或者称为次音信号;高于20kHz的信号称为超音频信号,或称为超声波信号。声音的频率范围如表5-1所示。

表 5-1 声音的频率范围

分 类	亚 音 频 波	音 频 波	超 音 波
频 率	小于 20Hz	20~20 000Hz	大于 20 000Hz

按照来源及作用,声音可分为人声、乐音和响音。人声包括人物的独白、对向、旁白、歌声、啼笑、感叹等;乐音也可称为音乐,是指人类通过相关乐器演奏出来的声音,如影视作品中的背景声音,一般起着渲染气氛的作用;响音是指除语言和音乐之外电影中所有声音的统称,如动作音响、自然音响、背景音响、机械音响、特殊音响。

5.2 音频处理设备

5.2.1 声音记录设备

最初声音信息的传播是瞬时性的,不能对声音进行存储和回放。直到爱迪生发明留声机,声音才得以记录和重放。爱迪生的留声机记录声音利用的是“声音是由振动产生的”这一基本原理。爱迪生的留声机从发明到现在,其设备外形和记录介质已有天壤之别,在记录形式和记录技术上也有所不同。下面通过几种不同的声音记录设备来简述声音记录技术的发展历史。

1. 机械留声机

最早用来记录声音的是机械留声机,是1877年由美国人爱迪生发明的。初期的留声机结构非常简单,只是在一个木盒中装上一只铜制的大喇叭,录、放音的声波都经过这只喇叭传递。之后,留声机不断改进,出现了爱氏(爱迪生)和贝氏(伯利纳,Emil Berliner)留声机,当然音质很差,频率仅为200~3000Hz,失真、噪声都较大,动态范围很小。1925年,美国的贝尔研究所开始运用电子管放大器和传声器进行电气灌片——电唱机。虽然机械式留声机质量很差,但由于在当时它是唯一的聲音记录设备,所以还是经历了50年之后才销声匿迹。

2. 钢丝录音机

世界上最早出现的钢丝录音机是1898年由丹麦科学家波尔森发明的,第一次用磁性记录的方式进行记录。它是把声音信号记录在钢丝上,最初的钢丝录音机采用无偏磁录音,失真度很大,到1907年之后才发明并出现用直流偏磁法进行录音,失真和灵敏度才大有改观,但信噪比仍然很差。直到1927年美国科学家卡尔逊发明了采用交流偏磁技术的专利才使音质大为

提高,很多 20 世纪 40 至 50 年代的优秀文艺节目都是利用这种录音机保留下来的。20 世纪 60 年代初还在使用这种机器。

3. 磁带录音机

磁带录音机属于磁性记录技术的再发展,是根据电磁感应定律,提出用永久剩磁录音的可能性,把声音记录在磁带上,接着再用磁带进行还原。1926 年,美国人开始用纸质带基制作成最初的磁带,后来进化成塑料带基涂敷型磁带,进而出现环形磁头、氧化铁粉磁带。

这时的录音机已基本定型为开盘式的,一直到 1949 年英国马可尼公司利用非线性磁头制成世界上第一台立体声录音机。立体声录音很快风靡全世界,许多录音机制造厂也就应运而生。

录音机由开盘式衍生出各种各样的盒式录音机,广播电台、电视台、电影厂的盒式专业录音机和民用录音机纷纷出现。由于音频技术的迅猛发展,在机型的繁衍、结构的改进、功能的扩展、性能的提高等诸多方面都取得了令人瞩目的进步。

4. 激光唱机

激光唱机是 1982 年由荷兰飞利浦公司和日本索尼公司联合开发推出的以数字方式记录的媒体,是一种用计算机控制的智能化高保真立体声音响设备。它采用先进的激光技术、数字信号处理技术、大规模集成电路技术以及自动控制、精密机械等先进技术,具有记录密度高、放音时间长、操作简单、选曲快、重放的声音层次分明及临场感强等优点,是目前最好的高质量音源设备。

激光唱机通常称为 CD 机,CD 是英文 Compact disc-Digital audio 的缩写,原意为“数字化精密型唱片及放唱系统”,该系统由激光唱片和激光唱机组成。

5. MP3 随身听

MP3 随身听的操作方法分为准备 MP3 音乐节目和播放两个过程。将 MP3 音乐文件由 PC 下载到 MP3 随身听并存储在闪存卡或微型硬盘内。播放过程就是 MP3 音乐文件的解码过程,多点控制单元(Multipoint Control Unit, MCU)利用 CPU(中央处理器)对 MP3 音乐文件进行解码运算,也有使用专用数字信号处理(Digital Signal Processing, DSP)来解码的,然后将数字音频信号经数-模转换器(Digital-to-Analog Converter, DAC)转换为模拟音频信号并用耳机来收听。

上述材料中显示了传统音频记录技术的演变历史。从记录介质上看,历经了石蜡(锡箔)记录、钢丝记录、磁带记录、激光数字记录;从技术手段上来看,经历了机械记录和磁性记录、激光记录;从外形上来看,录音设备由原来的开放式结构变成后来的封闭式设备(盒式)。

5.2.2 音频制作设备

在对声音进行处理的过程中,除对声音进行记录之外,还需要对声音进行一些其他方面的调整,如对声音进行音调的调节、多声音混合、高中低音的调整,还有诸如原始声波信号的拾取等。这就涉及音频制作设备。

1. 话筒

话筒(Microphone, 麦克风)的主要功能就是进行声音能量的收集。例如,在机械留声机中的铜制大喇叭除了完成还原声音(播放)这一功能,另一个功能就是在记录声音时进行声音能量的收集,这可以说是最早的话筒雏形。当出现磁性记录技术之后,话筒的功能就开始发生变化,除了完成声音的收集,还要完成声能向电能的转化(声音信号转化成电流信号),但是其

还原声音(播放)的功能已逐渐消失。

2. 音箱

音箱(Speaker,扬声器)的主要功能就是还原声音,将音频电流信号变换成声音信号,可以说是留声机中大喇叭另一功能的转化。

3. 模拟调音台

调音台在现代电台广播、舞台扩音、音响节目制作中是一种经常使用的设备。它具有多路输入,每路的声音信号可以单独被处理。例如,可放大,做高音、中音、低音的音质补偿,给输入的声音增加韵味,对该路声源做空间定位等;还可以进行各种声音的混合,且混合比例可调;拥有多种输出(包括左右立体声输出、编辑输出、混合单声输出、监听输出、录音输出及各种辅助输出等)。调音台在诸多音频系统中起着核心作用,它既能创作立体声、美化声音,又可以抑制噪声、控制音量,是声音艺术处理必不可少的一种设备。

5.3 音频的数字化

在传统音频处理技术中,通常处理的是模拟信号。一般情况下,模拟信号在时间或者空间维度上可以无限制地细分下去。模拟信号最大的特点就是,它是一种连续的不间断信号。如果用数学函数来表示,模拟信号的函数属于连续函数,在空间坐标轴上可以描出函数曲线上的无数个点。

在对音频模拟信号进行处理时,一般采用模拟的技术手段。例如,在记录音频信号时,是用无数个连续变化的磁场状态来记录,人们根本无法从中找到一个能代表声波元素的绝对磁场强度,而且每个点的磁场强度也不是单独存在的。在进行磁性信号记录之前,信号的传递是线性的,即便是在不同能量形式之间的转换(声能—电能—磁能)也是如此,信号各点之间的关系是不变的。电器元件是将连续的原始信号的变化形式原封不动地传递给下一个单元,这就是模拟的处理方式。

“原封不动”是指信号本身的连续性,在信号的强弱和纯净度上或许由于电子器件本身的频响特性而有一定的出入。例如,在记录时,会出现磁性材料的自身噪声难以完全去除,模拟电子器件信号传输过程中的噪声电流无法避免等问题,这是模拟处理方式的缺陷所在。

由于模拟音频处理技术对声音的处理会无法避免地引入噪声,难以最大限度地保持声音的原始效果,且存储介质的磁性变化会直接影响模拟音频的回放质量。因此,科学工作者开始探究数字音频技术,试图通过数字来保存声音,而且目前数字音频技术已经取得了良好的实际应用效果。那么,究竟什么是数字音频?数字音频技术解决了什么问题?数字音频技术中的关键是什么?

5.3.1 音频数字化过程

模拟音频信号在能量转换、传输和记录存储以及声音信号的重放过程中,其信号是连续不间断的。实际上,根据人耳的听觉特性,人们是无法区别间隔微秒级别以上声音的前后差别的,而且模拟音频处理设备在工艺上也无法保证在微秒级别上的前后声音频率响应特征完全不一致。

声音信号是时间和幅度上都连续的模拟信号。而计算机只认识0和1,或者说计算机只能处理一个个数据,尽管数据量可能是巨大的。所以,计算机处理声音的第一步是将声音数字

化,将模拟信号变为数字信号。

把模拟声音信号转换为数字化声音的过程称为声音的数字化,或称为模/数(A/D)变换。模拟音频在时间和幅度上都是连续变化的。数字音频在时间和幅度上都是离散、不连续的。数字音频是指用一连串二进制数据来保存的声音信号。这种声音信号在存储和电路传输及处理过程中,不再是连续的信号,而是离散的信号。关于离散的含义,可以这样去理解,例如某一数字音频信号中,数据A代表的是该信号中的某一时间点 a ,数据B是记录的时间点 b ,那么时间点 a 和时间点 b 之间可以分多少时间点,就已经固定,而不是无限制的。也就是说,在坐标轴上描述信号的波形和振幅时,模拟信号是用无限个点去描述,而数字信号是用有限个点去描述。

数字音频只是在存储和传输处理过程中采用离散的数据信号方式,而非全部的音频处理过程。因为在采集数字音频时的处理对象(音源信号)以及还原数字音频时的所得信号其实都还是模拟信号。数字信号与模拟信号相比较而言,具有处理技术简单、传输过程中无噪声及可多次无损复制等优点。音频处理倾向于采用数字音频技术,而且只需一台多媒体计算机和简单的配套设施,就可以组建起个人音频工作室。

模拟声音的数字化处理包括采样、量化和编码三个过程。连续时间的离散化通过采样来实现,连续幅度的离散化通过量化来实现。模拟声音的数字化过程如图5-1所示。

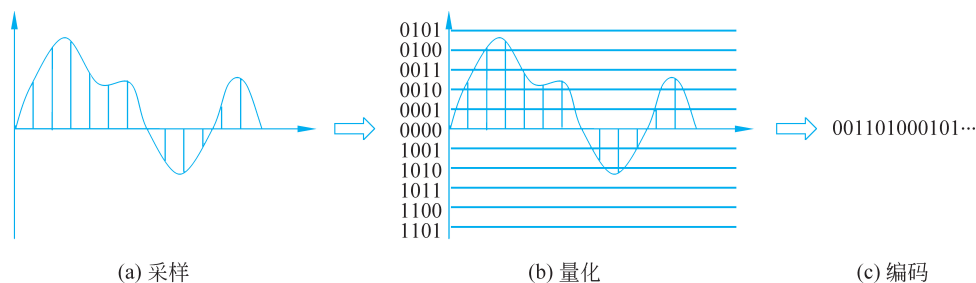


图 5-1 模拟声音信号的数字化过程

1. 采样

为实现A/D转换,需要把模拟声音信号波形进行分割,以转换成数字信号,这种方法称为采样(Sampling)。

采样的过程是每隔一个时间间隔在模拟声音的波形上取一个幅度值,把时间上的连续信号变成时间上的离散信号。该时间间隔称为采样周期,其倒数为采样频率。

采样频率是指计算机每秒采集多少个声音样本。采样频率越高,即采样的间隔时间越短,则在单位时间内计算机得到的声音样本数据就越多,对声音波形的表示也越精确。

采样频率的选择与声音信号本身的频率有关,根据奈奎斯特(Nyquist)理论,只有采样频率高于声音信号最高频率的两倍时,才能把数字信号表示的声音较好地还原为原来的声音。最常用的采样频率有11.025kHz、22.05kHz、44.1kHz等。

2. 量化

采样所得到的声波上的幅度值,即某一瞬间声波幅度的电压值,影响音量的高低,该值的大小需要用某种数字化的方法表示。通常把对声波波形幅度的数字化表示称为量化(Quantization)。

量化的过程是先将采样后的信号按整个声波的幅度划分成有限个区段的集合,把落入某个区段内的采样值归为一类,并赋予相同的量化值。采样信号的量化值采用二进制表示,表示

采样信号的幅度二进制的位数称量化位数。如果以 8b 为记录模式,则将其纵轴划分为 28 个量化等级,它的量化位数为 8。

在相同的采样频率之下,量化位数越高,声音的质量越好。同样,在相同量化位数的情况下,采样频率越高,声音效果也就越好。这就好比是量一个人的身高,若是以 mm(毫米)为单位来测量,会比以 cm(厘米)为单位量更加准确。

3. 编码

模拟信号经过采样和量化以后,形成一系列的离散信号——脉冲数字信号。这种脉冲数字信号可以用一定的方式进行编码,形成计算机内部运行的数据。所谓编码,就是按照一定的格式把经过采样和量化得到的离散数据记录下来,并在有效的数据中加入一些用于纠错同步和控制的数据。在数据回放时,可以根据所记录的纠错数据判别读出的声音数据是否有错,如果有错,可加以纠正。

5.3.2 数字音频的编码技术

数字声音的编码技术分为三类:波形编码、参数编码和混合编码。

1. 波形编码

波形编码是声音信号常用的编码方法,它直接对波形采样、量化和编码,算法简单,易于实现。而且,声音恢复时能保持原有的特点,因此被广泛应用。常用的波形编码方法有脉冲编码调制(Pulse Code Modulation, PCM)、差分脉冲编码调制(Differential Pulse Code Modulation, DPCM)和自适应差分编码调制(Adaptive Differential Pulse Code Modulation, ADPCM)等。

1) PCM

PCM 简称脉码调制,可以直接对声音信号做 A/D 转换,用一组二进制数字编码表示,得到的是未经压缩的声音数据。这是一种最常用、最简单的编码方法。

PCM 编码方法不需要复杂的信号处理技术就能实现瞬时的数据量化和还原,而且信噪比高。在解码后恢复的声音,只要采样频率足够高,量化位数足够多,就会有很好的质量。但是,这种对声音信号直接量化的方法编码数据量很大,需要很高的传输速率。

在多媒体个人计算机(Multimedia Personal Computer, MPC)中,声卡都具有 PCM 编码和解码的功能。激光唱盘(Compact Disc-Digital Audio, CD-DA)记录声音时就采用这种方法,存储未经压缩的数字声音信号。

PCM 编码是波形压缩编码的基础,波形压缩编码把 PCM 编码作为输入,并对其进行压缩。

2) DPCM

DPCM 编码是利用声音信号的相关性,通过只传输声音的预测值和样本值的差值来降低声音数据的编码率的一种方法。它采用预测编码技术,实现声音数据的压缩编码。

因为声音信号一般不会发生突然变化,相邻的语音采样值之间存在很大的相关性,从一个采样值到相邻的另一个采样值的差值要比样本值本身小得多。利用预测编码方法建立预测模型,通过预测器对未来的样本进行预测,然后对样本值与预测器得到的预测值之差进行量化和传输。由于这个差值的幅度远远小于样本值本身,需要较少的比特数来表示,这样可以降低数据的编码率,从而使编码数据得到压缩。

3) ADPCM

在实际使用中,由于输入信号的不稳定性,造成 DPCM 方法的信噪比大大降低。因此,在 DPCM 编码中加入自适应的方法,就形成了自适应差分编码调制(ADPCM)方案。所以,ADPCM 是对 DPCM 方法的改进,通过调整量化位数,对不同的频段设置不同的量化位数,可使数据得到进一步压缩。

ADPCM 压缩方案的压缩倍率可达 2~5 倍,信噪比高,性能优越,因此,多媒体计算机所获得的数字化的声音信息大都采用此压缩方法。MPC 的声卡也提供有 ADPCM 算法,如将 16 位的采样值压缩成 4 位,将 8 位的采样值压缩成 4 位、3 位或 2 位。

2. 参数编码

波形编码方法的编码率比较高,但可以获得较好的音质。除此之外,还有一类的编码方法是通过建立声音的产生模型,将声音信号以模型参数表示,再对参数进行编码,这种编码方法称为参数编码。声音重放时,再根据这些参数通过合成各种声音元码来产生声音。参数编码压缩倍数很高,但计算量大,而且保真度不高,合成声音的质量不如波形编码,所以,它适合于语音信号的编码。

3. 混合编码

将波形编码和参数编码的方法结合起来就称为混合编码。这是一种吸取波形编码和参数编码的优点进行综合编码的方法,它在降低数据率的同时,能够得到较高的声音质量。典型的混合编码方法有码激励线性预测(Code-Excited Linear Prediction, CELP)编码和多脉冲激励线性预测编码(Multi-Rate Linear Predictive Coding, MRLPC)等。

5.3.3 数字音频编码标准

当前编码技术发展的一个重要方向就是综合现有的编码技术,制定全球的统一标准,使信息管理系统具有普遍的互操作性,并确保未来的兼容性。

1. ITU-T G 系列声音压缩标准

国际电报电话咨询委员会(International Telegraph and Telephone Consultative Committee, CCITT)和国际标准化组织(ISO)先后提出一系列有关声音编码的建议,对语音信号压缩编码的审议在 CCITT 下设的第十五研究组进行,相应的建议为 G 系列,多由国际电信联盟(ITU)发表。

1) G.711

本建议公布于 1972 年,它给出了话音信号的编码的推荐特性。话音的采样率为 8kHz,每个样值采用 8 位二进制编码,推荐使用 A 律和 μ 律编码。本建议中分别给出了 A 律和 μ 律的定义,它是将 13 位的 PCM 按 A 律,14 位的 PCM 按 μ 律转换为 8 位编码。

2) G.721

本建议公布于 1984 年,1986 年做了进一步修订。采用自适应差值量化的算法对声音波形编码,数据率为 32kb/s,用于把 64kb/s 的 A 律或 μ 律的 PCM 编码转换成 32 kb/s 的 ADPCM 编码,实现对 PCM 信道的扩容。

G.721 和 G.711 标准都适用于 200~3400Hz 窄带话音信号,可用于公共电话网。

3) G.722

本建议公布于 1988 年,它是针对宽带语音制定的标准,给出了 50~7000Hz 声音编码系统的特性,可用于各种高质量语音。它的编码系统采用子带自适应差分脉冲编码(Subband

Adaptive Differential Pulse Code Modulation, SB-ADPCM) 技术, 整个频带分成高和低两个子带, 用 ADPCM 分别对每个子带进行编码。系统的比特率为 64kb/s, 所以称为 64kb/s(7kHz) 声音编码。

4) G.728

为了进一步降低数据速率, 实现低码率、低延时、高质量的目标, 在美国电话电报公司 (American Telephone and Telegraph Company, AT&T) 的贝尔实验室研究的 16kb/s 低延时码激励编码方案 (Low-Delay Code Excited Linear Prediction, LD-CELP) 的基础上, CCITT 于 1992 年和 1993 年分别公布了浮点和定点算法的 G.728 语音编码标准, 该算法编码延时小于 2ms。这个标准可用于可视电话、无绳电话、数字卫星通信、公共电话网、综合业务数字网 (Integrated Services Digital Network, ISDN)、数字电路倍增设备 (Digital Circuit Multiplication Equipment, DCME)、声音存储和传输系统、声音信息的记录和发布、地面数字移动雷达和分组化语音等。

2. MPEG 中的声音编码

MPEG (Moving Picture Experts Group, 动态图像专家组) 是国际标准化组织/国际电工委员会 (ISO/IEC) 于 1988 年成立的专门针对运动图像和语音压缩制定国际标准的组织, 其制定推荐了视频、音频、数据的压缩标准, 即 MPEG 标准。MPEG 标准主要有五个, 即 MPEG-1、MPEG-2、MPEG-4、MPEG-7 及 MPEG-21。

MPEG-1 声音压缩编码是国际上第一个高保真声音数据压缩的国际标准, 它分为三个层次。

(1) 层 1 (Layer 1): 编码简单, 用于数字盒式录音磁带。

(2) 层 2 (Layer 2): 算法复杂度中等, 用于数字音频广播 (Digital Audio Broadcasting, DAB) 和视频压缩光盘 (Video Compact Disc, VCD) 等。

(3) 层 3 (Layer 3): 编码复杂, 用于互联网上的高质量声音的传输。

MP3 是 MPEG Audio Layer-3 的简称, 是 MPEG-1 的衍生编码方案, 可以做到 12:1 的惊人压缩比并保持基本可听的音质。MP3 是目前最为普及的声音压缩格式, 各种与 MP3 相关的软件产品层出不穷, 而且更多的硬件产品也开始支持 MP3。随着网络的普及, MP3 被数以亿计的用户所接受。

3. AC-3 编码和解码

AC-3 (Audio Coding-3) 声音编码标准起源于由美国的杜比 (DOLBY) 公司推出的 DOLBY AC-1。AC-1 应用的编码技术是自适应增量调制 (Adaptive Delta Modulation, ADM), 它把 20kHz 的宽带立体声声信号编码成 512kb/s 的数据流。AC-1 曾在卫星电视和调频广播上得到广泛应用。1990 年, DOLBY 实验室推出了立体声编码标准 AC-2, 它采用类似修正离散余弦变换 (Modified Discrete Cosine Transform, MDCT) 的重叠窗口的快速傅里叶变换 (Fast Fourier Transform, FFT) 编码技术, 其数据率在 256kb/s 以下。AC-2 被应用在 PC 声卡和综合业务数字网等方面。

1992 年, DOLBY 实验室在 AC-2 的基础上, 又开发了 AC-3 的数字声音编码技术。AC-3 提供了五个声道, 20Hz~20kHz 的全通带频, 即正前方的左 (L)、中 (C) 和右 (R), 后边的两个独立的环绕声通道左后 (LS) 和右后 (RS)。AC-3 同时还提供了一个 100Hz 以下的超低音声道供用户选用, 以弥补低音的不足, 此声道仅为辅助而已, 故定为 0.1 声道, 所以 AC-3 被称为 5.1 声道。AC-3 将这六个声道进行数字编码, 并将它们压缩成一个通道, 而它的比特率仅为

320kb/s。

5.3.4 数字音频信息的质量与数据量

采样、量化和编码技术是声音数字化的关键技术。而采样频率、每个采样值的量化位数以及声音信息的声道数目,是影响数字化声音信息质量和存储量的三个重要因素。采样频率越高,量化位数越多,声道数目越多,声音的质量就越高,但存储量就越大。

1. 声音质量的评价

声音质量的评价是一个很困难的问题,也是一个值得研究的课题。目前,声音质量的度量有两种基本方法:一种是客观的质量度量;另一种是主观的质量度量。

1) 客观的质量度量

对声波的测量包括评价值的测量、声源的测量和音质的测量,其测量与分析工作是使用带计算机处理系统的高级声学测量仪器来完成的。度量声音客观质量的一个主要指标是信噪比(Signal to Noise Ratio, SNR)。信噪比是有用信号与噪声之比的简称,其单位是分贝(dB)。信噪比越大,声音质量越好。

2) 主观的质量度量

采用客观标准方法很难真正评定编码器的质量,在实际评价中,主观的质量度量比客观的质量度量更为恰当、合理。主观的质量度量通常是对某编码器输出的声音质量进行评价。例如,播放一段音乐或记录一段话,然后重放给一批实验者听,再由实验者进行综合评定,得出平均判分(Mean Opinion Score, MOS)。这种判分采用五分制,如表 5-2 所示,不同的 MOS 对应不同的质量级别和失真级别。

表 5-2 MOS 标准

MOS	质量级别	失真级别
5	优(Excellent)	未察觉
4	良(Good)	刚察觉但不可厌
3	中(Fair)	察觉及稍微可厌
2	差(Poor)	可厌但不令人反感
1	劣(Unacceptable)	极可厌(令人反感)

3) 常用的数字化声音技术指标及音质

常用的数字化声音技术指标及声音的质量有表 5-3 所列的几种。

表 5-3 常用的数字化声音技术指标及声音的质量表

采样频率 /kHz	量化位数 /b	每分钟数据量(无压缩)/MB		常用编码 方法	质量与应用
		单声道	双声道		
44.1	16	5.05	10.09	PCM	相当于 CD 质量,应用于超高保真质量要求
22.05	16	2.52	5.05	ADPCM	相当于收音质量,可应用于伴音及各种音响效果
	8	1.76	2.52	ADPCM	

续表

采样频率 /kHz	量化位数 /b	每分钟数据量(无压缩)/MB		常用编码 方法	质量与应用
		单声道	双声道		
11.025	16	1.76	2.52	ADPCM	相当于电话质量,可应用于伴音或解说词的录制
	8	0.63	1.26	ADPCM	

2. 声音信息的数据量

确定了数字声音的采样频率、量化位数和声道数就可以计算出声音信息的数据量,其计算公式为

$$S=R \times r \times N \times D / 8$$

其中, S 表示文件的大小,单位是 B; R 表示采样频率,单位是 Hz; r 表示量化位数,单位是 b; N 表示声道数; D 表示录音时间,单位是 s。

5.3.5 数字音频的文件格式

在多媒体计算机中,存储声音信息的文件格式有许多,主要有 WAV 格式、MP3 格式、WMA 格式、RealAudio 格式、MIDI 格式、QuickTime 格式、VQF 格式、DVD Audio 格式、MD 格式和 VOC 格式等。

1. WAV 格式

WAV 的全称是 Waveform Audio File Format,它是 Microsoft 公司开发的一种波形文件格式,采用 wav 作为扩展名,是如今计算机中最为常见的声音文件类型之一。它符合 RIFF 文件规范,用于保存 Windows 平台的音频信息资源,被 Windows 平台机器应用程序所广泛支持。WAV 格式存放的是未经压缩处理的音频数据,利用该格式记录的声音文件能够和原声基本一致,音频质量高,但由于体积很大(1min 的 CD 音质需要 10Mb),不适于在网络上传播。

2. MP3 格式

MP3 是一种以高保真为前提实现的高效压缩技术,压缩率为 10 : 1~12 : 1。

用 MP3 格式来存储,一般只有 WAV 格式文件的 1/10,而音质要次于 WAV 格式的声音文件。MP3 格式采用高压缩率编码方式,文件较小,所以 MP3 成为目前最为流行的一种音乐文件。

3. WMA 格式

WMA 的全称是 Windows Media Audio,它是 Microsoft 公司推出的一种音频格式。压缩率一般都可以达到 18 : 1。

4. RealAudio 格式

RealAudio 格式是一种流式音频文件格式,最大的特点就是可以实时传输音频信息,尤其在网速较慢的情况下,仍然可以较为流畅地传输数据,因此,RealAudio 主要适用于网上在线音乐欣赏。现在的 RealAudio 格式主要有 RA (RealAudio)、RM (RealMedia) 和 RMX (RealAudio Max) 三种。

5. MIDI 格式

MIDI 是 Music Instrument Digital Interface 的缩写,又称乐器数字接口,是数字音乐电子合成乐器的统一国际标准。它定义了计算机音乐程序、数字合成器及其他电子设备交换音乐信号的方式,规定了不同厂家的电子乐器与计算机连接的电缆和硬件及设备间数据传输的协