

完全手册系列丛书

MP3

完全手册



怀石工作室 编著



中国电力出版社

www.cepp.com.cn

455509

完全手册系列丛书

MP3 完全手册



怀石工作室 编著



00455509

中国电力出版社

内 容 提 要

MP3 是一个火热的名字。本书详细介绍了 MP3 的原理、常用技术,介绍了 MP3 软件的概况。随后介绍了比较常见的 MP3 播放软件、CD 转录软件、整合性软件,MP3 随身听的功能、技术特点、使用方法和选购指南。在第八章中,着重介绍了制作 MP3 的技巧和相关软件的使用方法。本书的最后是有关 MP3 的一些资料,包括 MP3 站点信息等。

本书适合对 MP3 感兴趣的读者和 MP3 迷阅读。

图书在版编目(CIP)数据

MP3 完全手册/怀石工作室 编著. —北京:中国电力出版社,2000.1

(完全手册系列丛书)

ISBN 7-5083-0168-4

I. M… II. 怀… III. 音乐-数字信号-压缩-应用软件, MP3
IV. TP391.4

中国版本图书馆 CIP 数据核字(1999)第 63766 号

中国电力出版社出版、发行

(北京三里河路 6 号 100044 <http://www.cepp.com.cn>)

三河市实验小学印刷厂印刷

各地新华书店经售

*

2000 年 1 月第一版 2000 年 1 月北京第一次印刷

787 毫米×1092 毫米 16 开本 17 印张 381 千字

定价 25.00 元

版 权 所 有 翻 印 必 究

(本书如有印装质量问题,我社发行部负责退换)

前 言

自从人类将数字技术引入到音响领域以来，已经有四分之一个世纪。但真正大发展还是最近十余年的事情。但正如人类在其它技术领域取得的成就一样，经过初期的实践和积累阶段后，数字音响技术呈极大的加速度向前发展。数字音响的应用，使音质产生了质的飞跃。近年来，随着互联网的飞速发展，音频压缩技术得到广泛的应用。到目前为止，这一领域中，最为广大用户所接受的就是 MP3 了。

从 MP3 诞生之日起，它就备受关注。这些关注中，有惊奇，有赞叹，有期望，还有敌视。是什么使 MP3 具有了如此“魔力”？MP3 作为当前数字音频压缩技术的典型代表，了解和掌握其基本知识，对所有的电脑“玩家”都是非常有意义的。这不仅是知识面的拓宽，也不仅是多了一种娱乐的方式，从更广阔的范围来看，技术发展具有的承接性，使得典型技术最容易成为新技术发展的起点。因此，了解 MP3 也是把握数字音频压缩技术发展脉搏的重要途径。

现在的电脑报刊中，MP3 已成为一个出现频率级高的词，但仅仅是篇幅的限制，就使得它们不可能对 MP3 做出全面细致的介绍，这也成为写作本书的重要原因之一。

本书从内容上，可分为三个部分：

第一章从 MP3 的概念讲起，较为详细地介绍了其中的原理、常用技术；第二章介绍了 MP3 软件的概况。这两章构成本书的基础篇。

从第三章到第七章，依次介绍了比较常见的 MP3 播放软件、CD 转录软件、MP3 编码软件、整合性软件和 MP3 随身听的功能、技术特点和使用方法，从而构成了本书的实用篇。

在第八章中，着重介绍的一些制作 MP3 的技巧和相关软件的使用方法，可以作为提高篇。

本书的最后，有三个附录，提供了一些有关 MP3 的资料，包括站点信息，供有兴趣和有条件的读者参考。

对没有接触过 MP3 的读者来说，可以从本书获得入门所需的相关知识；对已经具有的一定 MP3 知识的读者，也可以从书中更深入地了解 MP3 技术的一些细节和软件的使用技巧，从而得到进一步的提高。

由于时间和编者水平所限，书中肯定有诸多不足甚至错误之处，希望读者和同行特别是广大 MP3 迷们能提出宝贵意见。

编者

1999 年 11 月 23 日

目 录

第一章	走近 MP3	1
1.1	什么是 MP3	1
1.2	MP3 与 MPEG	1
1.3	MP3 原理浅析	4
1.4	MP3 的发展史	12
1.5	MP3 涉及的法律问题	15
1.6	MP3 的发展趋势与面临的挑战	20
第二章	MP3 软件总览	25
2.1	MP3 软件的分类	25
2.2	获取 MP3 软件的途径	33
第三章	MP3 播放器	35
3.1	Winamp	35
3.2	K-jofol	67
3.3	Wplay Pro	82
3.4	FreeAmp	94
第四章	转录 CD 音轨	104
4.1	转换为 WAV 格式	104
4.2	直接转录为 MP3	127
第五章	MP3 压缩软件	140
5.1	DOS 平台下的编码软件	140
5.2	Windows 平台下的压缩软件	144
第六章	三合一集成软件	183
6.1	MusicMatch Jukebox	183
6.2	Real Jukebox	199
6.3	AudioCatalyst	209

第七章	MP3 随身听	221
7.1	Diamond Rio 随身听	222
7.2	选购 MP3 随身听	226
7.3	MP3 随身听市场展望	230
第八章	制作高质量的 MP3 作品	231
8.1	抓取高质量的 CD 音轨	231
8.2	修剪歌首和歌尾的静音	238
8.3	音量等化 (Normalize)	244
附录 1	MP3 文件头、帧头、ID3 格式和歌曲类型清单	247
附录 2	XingMP3 Encoder 和 X3Enc 支持的编码的模式	252
附录 3	MP3 站点大观	254

第一章 走近 MP3

相信每个接触过计算机的朋友都听说过 MP3。近几年来，MP3 随着 Internet 的普及，几乎成为每一台多媒体电脑的“座上宾”，“MP3”也一度成为几大搜索引擎上被检索频率最高的词。这可乐坏了音乐迷，也急坏、气坏了音乐出版商，真是几家欢喜，几家愁。那么，MP3 到底为何物？这股 MP3 “旋风”从何而起，又为何刮得如此强劲呢？MP3 将向何处去？别急，我们这就走近 MP3。

1.1 什么是 MP3

MP3 的全称为 MPEG Audio Layer3，是一种强大的计算机音频编码方案。它以较大的压缩比（通常为 1: 10 至 1: 12）将音频文件转换成较小的，扩展名为 .mp3 的文件。同时基本保持原文件的音质。MP3 使用的压缩算法是一种有损压缩方法，并不是保留原文件的所有信息，会产生一定的失真，但人耳不易察觉。MP3 是 ISO/MPEG 规范的一部分。它在现行的国际标准 MPEG-1（ISO/IEC 11172-3）和 MPEG-2（ISO/IEC 13818-3）的音频部分中都有定义，目前 MP3 绝大部分使用的是 MPEG-1 标准。在这里，请注意不要将 MPEG Layer3 称为 MPEG-3，因为存在 MPEG-1 和 MPEG-2 标准，而 MP3 是包含在上述两个标准中的，将其称为 MPEG-3 容易让人误解为 MPEG-3 标准。

1.2 MP3 与 MPEG

说到 MP3，我们不能不提一下 MPEG 标准。相信绝大多数电脑用户都不会对 MPEG 感到陌生。MPEG 是 Moving Picture Experts Group（动态图像专家组）的缩写。MPEG 受国际标准化组织（International Standards Organization, ISO）和国际电子技术协会（International Electro-Technical Commission, IEC）联合领导。它制定的 MPEG 标准（也就是 ISO/MPEG 标准）被广泛应用于各种多媒体中。VCD 和 DVD 就是用 MPEG 编码压缩的动态视频流文件。正是因为 MPEG 编码的高压缩率的特点，原本需要很大空间保存的音频、视频数据，在一张普通光盘上即可容纳。一张 VCD 光盘可保存 74 分钟的动态画面，而且其表现画质可达到 VHS 录像带的水平，音质也可达到或接近 CD 的标准。于是，一种自然而然的想法就产生了：如果将 MPEG 的编码技术用来压缩纯音频文件，那么又会

达到什么惊人的效果呢？从字面意思，我们不难推想，MPEG Layer 3 是 MPEG 的声音压缩标准之一，这就是 MP3 的诞生。

ISO/MPEG 标准描述了使用高性能感知编码方案的音频（Perceptual Coding Schemes）压缩。它一直在不断更新，新增更强的更有效率的编码方法，以满足人们对“质高量小”的追求。现在已经形成了一个包含上面提到的 MPEG Layer 1、Layer2 和 Layer3 三个音频编码方案的家族。压缩增益（Compression Gain，每位的音质）、编码复杂度、编码解码器所需的延迟从 Layer1 到 Layer3 递增。各方案的压缩率如表 1-1 所示。表中的“层数”表示不同的编码方案，可以理解为版本。

表1-1 ISO/MPEG 标准中各层的压缩率

层 数	常用压缩率
MPEG Layer 1	1: 4
MPEG Layer 2	1: 6 至 1: 8
MPEG Layer 3	1: 10 至 1: 12

从表 1-1 可知，MPEG Layer3 的压缩率可达 1: 10 至 1: 12。这个压缩比例可以说是很大了。在通常情况下，1M 的 MP3 文件可播放 1 分钟，也就是说一首 n 兆的 MP3 歌曲，你足足可欣赏 n 分钟以上。而一分钟的 CD 或 CD 音质的 WAV 文件（44100Hz，16bit，双声道，60 秒）要占用 10 兆左右的存储空间。这样算来，一张容量为 650 兆的 MP3 光盘，播放的时间应在 10 小时以上。而同样容量的一张 CD，播放时间在 70 分钟左右，MP3 在这一点上的优势是 CD 万难比拟的。

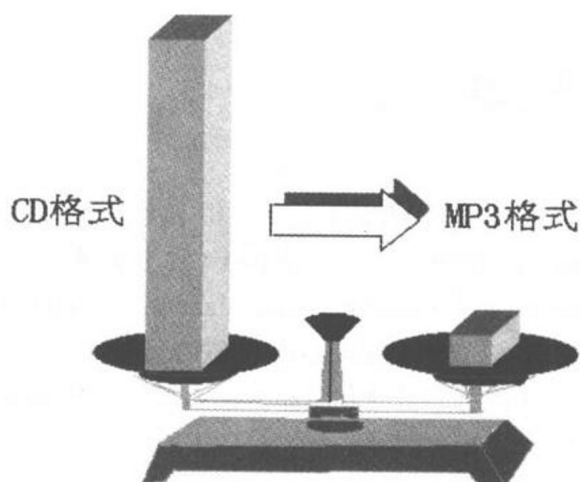


图1-1 MP3 拥有比 CD 高得多的信息密度

在 MPEG 标准中，所有的层使用相同的基本结构。三种编码方案通过继承的方式实

现兼容。也就是说，Layer N 的解码器可以对用 Layer N 和所有层数低于 N 的编码器生成的位流进行解码。如：Layer 3 解码器可接受 Layer1、Layer2、Layer3 编码器生成的文件，而 Layer2 解码器只接受 Layer1 和 Layer2 编码器生成的文件。对每一层，MPEG 标准指定了位流的格式和解码器。为进一步的发展，没有指定编码器，但是在说明中给出了编码器的示例。

MPEG 标准中还给出了各层的编码、解码器延迟的理论数值，如表 1-2 所示。

表1-2 MPEG 标准中给出的编码、解码器的理论延迟时间

层 数	理论延迟时间
MPEG Layer 1	19ms (<50ms)
MPEG Layer 2	35ms (<100ms)
MPEG Layer 3	59ms (<150ms)

表中只是理论上的延迟时间。实际数值明显高于这些理论值。由于延迟时间依赖于具体实现方法，所以预测这一数值是十分困难的。表中括号中的数值只是粗略的近似值——真正的编码、解码器延迟有可能比这些值还要高。因此，有些应用程序有可能受到延迟的影响，如对远程报表部件的反馈链接；而其他应用程序并不太关心这种延迟。

MP3 的压缩率如此大，难免让人担心它的效果。毫无疑问，MP3 作为一种数字压缩格式的文件，是必然要失真的，问题在于这种失真在多大程度上能被人耳区分出来。MP3 在这方面同样非常出色。实际上，如果我们同时用 MP3 和 CD 格式播放同一首歌曲，我们很难分辨出，哪个出自 MP3，哪个出自 CD。当然，这与使用的设备有一定关系。除非你是发烧级的音响迷，或是专业的音频工程师，否则配备可将 MP3 与 CD 音质区别出来的设备是没有必要的。

MP3 是一种有损压缩，也就是说在压缩时要过滤掉一些信号，不过这些信号多是人耳不易分辨的。这种方法类似于 JPG 图像格式对细节的处理。对细节的过滤应保持均匀过渡，即保持连续性，并限定在一定范围内。因为，人的耳朵和眼睛对渐变的信号的灵敏度低于对突变的信号的灵敏度；同时细节丢失得越多，失真越大。

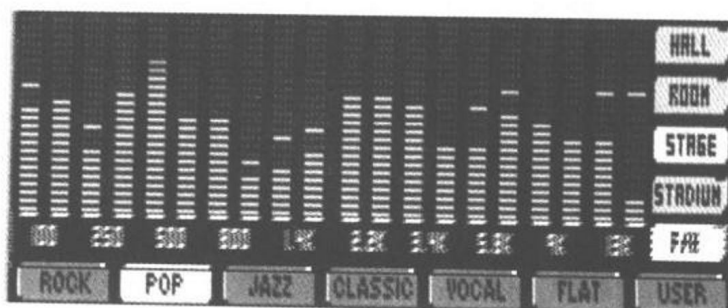


图1-2 MP3 的音频采样

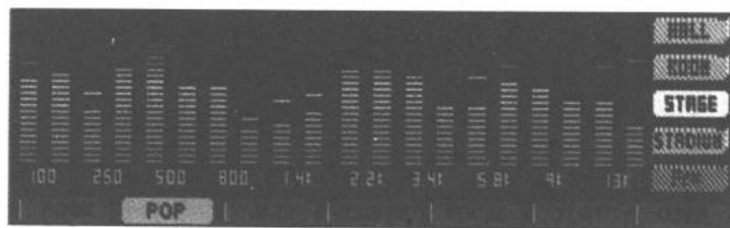


图1-3 CD 的音频采样

图 1-2 和图 1-3 分别是同一首歌曲分别以 MP3 格式和 CD 格式播放时，在同一点上的音频采样。容易看出，MP3 和 CD 在各个频率上，信号强度分布基本是一致的。不同的是，CD 在高频上的信号要强于 MP3，这说明 MP3 过滤掉一部分高频信号。稍后我们将详细介绍 MP3 编码的原理。

仅凭高密度、高音质这两件“法宝”，MP3 就足以成为 CD 的杀手。这正是令音乐出版商们头疼的地方，也是 MP3 迅速崛起的原因之所在。

1.3 MP3 原理浅析

随着计算机技术的快速发展，数字化音乐的发展也呈现加速发展的趋势。新技术和新名词层出不穷，常常使我们无所适从。我们无力跟踪、捕捉全部的新鲜事物，但这并不意味着我们只能在它们突然出现时，显得手足无措，任其摆布或人云亦云。我们在这里再次强调原理，是基于这样一种认识——无论是多么高新的技术，总是在其基础领域范围内产生、发展，并且在其发展过程中总是有一定的承继性。因此，只要了解了基本的概念和原理，很多新技术就变的容易理解，也就能抓住技术发展的脉络。

前面两节中，我们已经提到过，MP3 使用的是有损压缩方法。那么 MP3 为什么在压缩率如此之高的情况下，还能保持近乎 CD 的音质呢？这一节，我们就来讨论这个问题。我们还是先从一些基本概念入手。

📖 MPEG 音频压缩基础知识

“音频压缩(Audio Compression)”、“音频编码(Audio Encoding)”和“音频解码(Audio Decoding)”这些术语，很容易使人混淆，希望通过本节的讲解，能使你对它们有大体的了解。

📖 音频压缩的目的

前面已经提到，在音频压缩出现以前，高质量的数字音频数据要占用很大的硬盘空间或在传输时占用很大的带宽。

让我们再来看一个例子。你想对你喜爱的长为 4 分钟的歌曲采样，并将其存储在硬盘上。因为你想得到 CD 一样的音质，抽样率为 44.1kHz，立体声，每次抽样数据为 16 位。

44100 Hz 意味着每秒种，你需要从声卡或输入文件接收 44100 个值。将这个值乘以 2，因为你使用两个声道。因为每次采样的数据是 16 位即 2 个字节，所以要将上一步中得到的值再乘以 2。这样，这首歌所占的空间为：

$44100 \text{ 采样率} \times 2 \text{ 声道} \times 2 \text{ 字节/采样} \times 60 \text{ s/min} \times 4 \text{ min} \approx 40.4 \text{ MByte}$

如果你想从 Internet 上下载这首歌，假定你的“猫”的平均传输率为 28.8kbps，在理想的情况下，下载所需的时间为：

$40400000 \text{ Byte} \times 8 \text{ bit/Byte} / (28800 \text{ bps} \times 60 \text{ s/min}) \approx 187 \text{ 分钟}$

这仅仅是下载一首歌，上网费够买一张正版 CD 了！

在这一背景下，数字音频编码或者说是数字音频压缩，就成为减少音频数据的存储空间占用量或传输带宽的重要技术。现在的感知音频编码技术，比如 MPEG Layer-3 或 MPEG-2 AAC，利用人耳对声音感知的特性，在达到较高的压缩率的同时，可以使音质的损失不被或很少被察觉。

因此，这些方案就成为高音质、低数据流量应用程序使用的关键技术，如：CD-ROM 游戏的音轨、使用电晶体的声音存储器、Internet 音频和数字音频广播系统等等。

说明：

MPEG-2 AAC 是 MPEG-2 Advanced Audio Coding (MPEG-2 高级音频编码) 的缩写。MPEG-2 AAC 是 Fraunhofer IIS (德国的一家集成研究发展机构) 这样的研究机构同诸如 AT&T、Sony 和 Dolby 这样的大公司合作的结晶。MPEG 于 1997 年正式确认其为国际标准。

发展 AAC 的推动力量是对高效的环绕立体声信号编码方法的需求，如今天已在电影元中使用的 5 声道信号 (左、右、中、左环绕和右环绕)。在 MPEG-2 中，早就有针对这些信号的算法。但是由于技术和历史的原因，并没有得到最佳的效率。因此，AAC 指定的目标是能显著地减少所需的数据流量。

MPEG-2 AAC 是继空前成功的 ISO/MPEG Audio Layer-3 后，MPEG 的又一大贡献。国际合作的发展和从 Layer-3 中培养的洞察力为这种新的编码方法铺平了道路，这是没有先例的。高编码增益和高度的灵活性的恰当结合开辟了应用程序的广阔天地。采样率可在 8kHz 到 96kHz 之间，使用的声道数可在 1 到 48 之间，这为将来的发展留下了充分的余地。与 MPEG-2 Layer-2 这样的有名的编码方法相比，它有可能使数据流量下降一半，而主观音质却不受影响。

像所有的感知听觉编码方案一样，MPEG-2 AAC 基本上也是利用人耳的信号屏蔽特性来减少数据量。为做到这一点，量化噪音被分成频带。通过这种方式，它就被正常的信号所屏蔽，也就是说它变得听不见了。尽管这种编码方法的基本结构同它的“前辈”没有什么区别，但在细节上还是有许多值得注意的新特性。MPEG-2 AAC 同它的“前辈”

ISO/MPEG Audio Layer-3 有如下几点重要的不同之处:

滤波组: ISO/MPEG Audio Layer-3 使用的是混合滤波组,这主要是考虑到兼容性,但是在结构上存在一定的脆弱性。MPEG-2 AAC 使用的是简单的改进离散余弦变换 (Modified Discrete Cosine Transform, MDCT)。加上增加后的窗口长度 (每次变换为 2048 线而不是 1152 线), MDCT 的性能比以前的编码方法使用的滤波组都要出色。

临声噪声整形 TNS: TNS 是时间/频率编码方案中的新技术。它通过在频率域预测的方法,实时地将量化噪声整形。特别是语音信号通过 TNS 处理后,可获得很大改良。

预测: 预测是通常建立在语音编码系统领域内的一种技术。它得益于某些类型的音频信号的易预见性。

量化: 因为允许精密的量化分解控制,所以可以更加有效地利用给定的数据流量。

位流格式: 需要传输的信息经过平均信息量编码以使冗余信息尽可能地少。这些编码方法的优化和灵活的位流结构使进一步提高编码的效率成为可能。

音频压缩的两个组成部分

音频压缩实际上包含两个组成部分。第一部分称为编码,它将存在于 WAV 文件中的数字音频数据转换成高压缩率的形式,这种形式被称为位流或编码后的音频数据。为使声卡能够播放这种位流,还需要第二个部分,称为解码。解码过程接受位流并将其重建到 WAV 文件中。

音频压缩是如何工作的

最高的编码效率是利用特定的算法实现的。这些算法利用了频率域中的信号冗余和无关联性。频率域是建立在人类听力系统的模型上的。

所有的编码器使用相同的基本结构。编码方案可称为“感知噪声整形 (Perceptual Noise Shapping)”或“感知分波段/变换编码 (Perceptual Subband/Transform Coding)”。编码器通过计算滤波器组 (变换),来分析音频信号的谱线成分,并使用心理声学模型,来估算刚好能被察觉到的噪声水平。在量化和编码阶段,编码器通过符合数据流量和屏蔽两者要求的方式,试着分配有效的数据位编号。

同编码器相比,解码器要简单得多。它的任务只有一个,就是从编码后的谱线成分中,提取出声音信号。

压缩比例、数据流量和音质

直到现在,我们还没有明确指出,在编码和解码以后,你得到的并不是相同的文件。就是说,所有多余的信息 (更准确地说是声音信号的冗余和不关联的部分) 都已被排除在外了。重建的 WAV 文件有别于原来的 WAV 文件,但是听起来相差并不多,差别主要取决于对它压缩到了何种程度。

因为压缩率在某种程度上是一个难处理的尺度，所以当说到压缩的力度时，多使用数据流量这一术语。数据流量表明了每秒钟音频数据要消耗的平均 bit 数。通常的单位是 kbps，即千 bit 每秒，或 1000bit/s。

从 CD 中得到的数字音频信号的数据流量是 1411.2kbps。而使用 MPEG-2 AAC，得到的 CD 音质的信号的数据流量可达到 96 kbps。

📖 MPEG-1 与 MPEG-2

MPEG-1 和 MPEG-2 使用同一个音频编码解码族——Layer 1、Layer 2 和 Layer 3。MPEG-2 的新的音频一个特点是采用“低采样率扩展 (Low Sample Rate Extension)”，而将数据流量很低的应用程序和限定带宽的请求结合了起来。新的采样频率是 16、22.05 或 24KHz，数据流量向下扩展到 8kbps；另一特点是“多通道扩展 (Multichannel Extension)”，将主声道增加为 5 个，分别是左、中、右、左环绕和右环绕，还为子低音扬声器信号增加了一个可选的特殊“低频增强 (LFE)”声道，从而解决了环绕立体声应用程序的问题。另外，“多语言扩展”特性允许将加入的声道数提高到 7 个。

除了在低采样率方面以外，MPEG-1 和 MPEG-2 在其他方面是兼容的。很显然，MPEG-1 的解码器是不能处理新的“半”采样率的。在制定 MPEG-2 标准期间，兼容性问题一度成为最主要的话题。最主要的思想是使用与 MPEG-1 标准相同的基本位流格式。即象以前一样，主数据域带两个音频信号（称为 L0 和 R0）。而辅助数据域携带多声道扩展信息。在继续深入讨论之前，先介绍两个术语：“向前兼容”和“向后兼容”。“向前兼容”意味着 MPEG-2 标准的解码器必须接受任何 MPEG-1 标准的音频位流，这些位流代表着单声道或双声道；“向后兼容”意味着使用 MPEG-1 标准的解码器必须能够对使用 MPEG-2 标准的位流的主数据域（L0 和 R0）进行解码。在将环绕的音频信息转换入到 L0 和 R0 时，可能会使用矩阵变换的方法：

$$L0 = \text{左声道信息} + a \times \text{中央声道信息} + b \times \text{左环绕声道信息}$$

$$R0 = \text{右声道信息} + a \times \text{中央声道信息} + b \times \text{右环绕声道信息}$$

因此，使用 MPEG-1 标准的解码器可以再现全部 5 个声道的信息的完整的降阶矩阵。MPEG-2 解码器使用多声道扩展信息（多出的 3 个音频信号）来重建 5 个环绕声道。

说明：

“半”采样率是 Fraunhofer (Winplay3 的开发者) 编码解码器中使用的一种非 ISO 标准的扩展方法，用于提升低数据流量（比如说 8kbps）时的性能。MPEG-2 标准为实现低采样率扩展，允许数据流量低到 8kbps。在这样低的数据流量下，有用的音频带宽必须受到限制，比如 3kHz。因此实际的采样率可以降低，比如用 8kHz。采样率越低，频率分辨能力越好，时间分辨能力越差，位流格式中控制信息和有效载荷间的比例越好。MPEG-2 规定的最低采样率为 16kHz，在此基础上，可做进一步的扩展，比如将 MPEG-2 标准的采

样率除以 2，得到新的采样率为 8、11.025 和 12kHz。MP3 在数据流量 8kbps，采样率 8kHz 或数据流量 16kbps，采样率 11kHz 的情况下，比在 8 或 16kbps，采样率 16kHz 的情况下性能有明显提高。

MPEG Audio Layer-X 家族

在进一步了解了有关 MPEG-1 和 MPEG-2 的知识后，让我们来看一看 MPEG Audio Layer 1、Layer 2 和 Layer 3 三兄弟间的关系。前边已经提到过，所有的“层”使用相同的基本结构。

MPEG Layer-1、Layer-2 和 Layer-3 三个“层”主要有如下相同之处：

所有的“层”使用相同的分析滤波器组（有 32 个分波段的多相）。Layer-3 新加了 MDCT 变换，以增加频率分辨能力。

在位流中，所有的“层”使用相同的“头信息”，以支持标准中分层的结构。

各“层”对位错误有相似的敏感度。它们使用一种位流结构，其中包含的一些部分对位错误相对敏感，这些部分有“头标”、“位分配”，“比例因子”和“分区信息”。还有一些部分对错误不甚敏感，如“谱线成分数据”。

各“层”都支持在音频数据流中插入编程相关信息（即辅助数据）。

所有的“层”均可使用 32，44.1 或 48kHz 采样频率。

各“层”使用的数据流量均在 32kbps 以上。表 1-3 给出了各“层”可以使用的数据流量的范围；表 1-4 给出了各“层”常用的数据流量。

表1-3 各种方案可以使用的数据流量范围

层 数	可用数据流量
MPEG Layer-1	32kbps 至 448 kbps
MPEG Layer-2	32 kbps 至 384 kbps
MPEG Layer-3	32 kbps 至 320 kbps

表1-4 各种方案通常使用的数据流量

层 数	常用数据流量
MPEG Layer-1	384kbps
MPEG Layer-2	256kbps 至 192kbps
MPEG Layer-3	128kbps 至 112kbps

每一种方案都会扩展本层和低于本层的方案的特性。最简单的格式是 Layer-1。它主要是为数字压缩磁带 (Digital Compact Cassette——DCC) 设计的, 使用 384kbps 的数据流量, 称为“PASC”。Layer-2 在设计时, 在复杂性和性能间做了权衡。当数据流量下降到 192kbps 时, 还能保持很好的音质。低于 192kbps, 音质就要受到影响。Layer-3 从一开始就是为低数据流量设计的。它添加了一系列相对 Layer-2 的“高级特性”; 频率分辨能力是 Layer-2 的 18 倍, 这样 Layer-3 编码器可以使量化噪音更好地适应屏蔽阈值 (Masking Threshold)。在三种方案中, 只有 Layer-3 使用了平均信息量编码 (Entropy Coding), 这与 MPEG 视频类似, 可以更好地减少冗余信息; Layer-3 还是唯一使用位存储器 (类似 MPEG 视频) 的方案, 可以在临界时刻减少人为影响; Layer-3 还可以使用更高级的联合立体声编码方法。

通过使用立体声效果和限制带宽, 编码方案可以在较低的数据流量上得到很不错的效果。MPEG Layer-3 是编码家族中最强大的。就给定的音质水平, 它需要的数据流量最小, 或者说, 就给定的数据流量, 它可以得到最好的音质。

📖 关于音质

前面我们总提起音质。声音的品质反映到人的头脑中, 是一个客观见之于主观的过程。同样一首歌, 在不同的人听来, 音质可能是不同的。因此, 评价音质的标准是很难制定的。但要制定国际统一的音频处理标准, 音质的问题是不可回避的。音质的好坏到底是如何确定的呢?

今天, 唯一可行的方法是倾听测试法。在制定 ISO-MPEG 的过程中, 进行了大量的国际倾听测试, 这些测试都是由受过专门训练的人员进行的。所有这些测试中都使用“三元刺激 (Triple Stimulus)”和“隐藏参照 (Hidden Reference)”的方法, 并且使用“CCIR 损伤比例”确定音质。倾听的顺序是“ABC”, 其中 A=初始, BC=以随机序列形式给出的初始/编码信号对。听者必须用 1.0 到 5.0 之间的数字估计 B 和 C 的值。这些值的含义是: 5.0=透明的 (这应当是初始信号), 4.0=可察觉的, 但听起来, 并不会使人不快; 3.0=听上去使人稍有不快; 2.0=使人不快; 1.0=难以忍受。倾听测试是一项昂贵的工作。但目前还难以找到替代的方法, 以后或许会有所改观。如果使用感知编码解码器确定音质, 所有的传统音质参数, 像信号对噪声的比例、总谐波失真和带宽, 都会变得毫无用处, 因为任何一个编码解码器都可能引入噪声和失真, 只要它们不影响到感觉到的音质。因此倾听测试是必要的, 并且如果经过精心的准备和执行, 可以得到更可靠的结论。

📖 MPEG Layer-3 的压缩原理

“量小、质高”是 MP3 最大的特点, 下面就压缩频率及输出品质做一下比较, 如表 1-5 所示。

表1-5 MP3 的压缩率及声音品质

声音品质	声音频宽	模 式	数据流量	压缩比例
电 话 声	2.5kHz	单声道	8kbps	1:96
短波无线电	4.5kHz	单声道	16kbps	1:48
AM 调幅广播	7.5kHz	单声道	32kbps	1:24
FM 调频广播	11kHz	立体声	56 至 64kbps	1:24 至 26
近 似 CD	15kHz	立体声	96kbps	1:16
CD 音 质	大于 15kHz	立体声	112 至 128kbps	1:12 至 14

前面已经提到过，MP3 使用的“感知音频编码技术（Perceptual Audio Coding）”。这是一种失真算法。简单说来，就是裁剪掉人耳听不到的和不易听到的声音信号，以节省存储空间。

大家知道，正常人能听到的声音频率范围是 20Hz 至 20kHz。低于 20Hz 的次声波和高于 20kHz 的超声波人耳无法做出反应。但这并不意味着我们的录制设备不会记录这些信号。显然，这些信号对人类是无用的。从设备的方面讲，音响器材对 20Hz 到 20kHz 这一宽范围内的频率的响应程度也不相同。一般说来是中间好，两头差。这在客观上也为裁掉高频和低频的信号提供了依据。从声音的强度上说，人的听力系统具有非常优越的性能。其动态范围超过 96 分贝。人既可以听到针掉到地上的声音，也可以听到波音 747 飞机的轰鸣。但当我们站在飞机场听着波音 747 的轰鸣时，还能分辨出扣子掉在地上的声音吗？不可能。人的听力系统适应声音的动态变化，人们对这种适应及屏蔽特性音质研究后得出对声音压缩非常有用的理论。人们很早以前就知道利用这种特性来为磁带录音降低噪声了（如：当没有音乐声时，空白磁带的嘶嘶声是很容易听到的，而当音乐信号电平很高时，嘶嘶声就很难听到）。当声音较强时，对较弱的声音信号产生屏蔽效应。在阈值曲线下的噪音或信号较弱的声音无法被人耳听到。在较强信号出现时，允许通过更多的信号。在此时增加被量化过的弱信号数据（使用无用的位来携带更多的信息）可以达到一定程度的压缩的目的。通常情况下，MP3 压缩器将原始声音通过 FFT（快速傅立叶变换）变换到频率域，然后通过一定的算法算出何种频率声音可以携带更多的信息。这就形成了“有选择的差变”。而在还原时解码器所需要做的仅仅是将其从频域再变换回来。

说明：

动态范围是声音系统可以分辨的声音的强度范围。通常的音响设备由于动态范围的限制，多会造成一定的失真。例如：在自然界中，雷声要比雨点儿的声量大几万倍。但从音箱中放出来，差别就没这么大。雷声被减小了，而雨声被放大了。这就是动态范围引起的失真。在这方面，人耳具有不可比拟的优越性。

阈（值）是一种逻辑运算符，它具有如下特性：如果 P、Q、R...都是命题，那么若

至少有 N 个命题为真，则 P 、 Q 、 R ... 的阈值为真；若小于 N 个命题是真，则 P 、 Q 、 R ... 的阈值为假。其中 N 为指定的非负整数，称为阈条件。上述阈值曲线是建立在人的听力系统上的模型，进行阈运算的结果。

为了更好地说明 MP3 采用的压缩方法，还是让我们来看一个具体的例子。

图 1-4 中所显示的是一段 2.5 秒钟的音乐频谱，X 轴是音乐的时间轴（0 至 2.5 秒钟），Y 轴则是音乐的频率（0 至 20000Hz）。图的上半部是左声道，下半部是右声道。图中音乐频谱是直接从音乐 CD 上获取的，并未作任何处理。

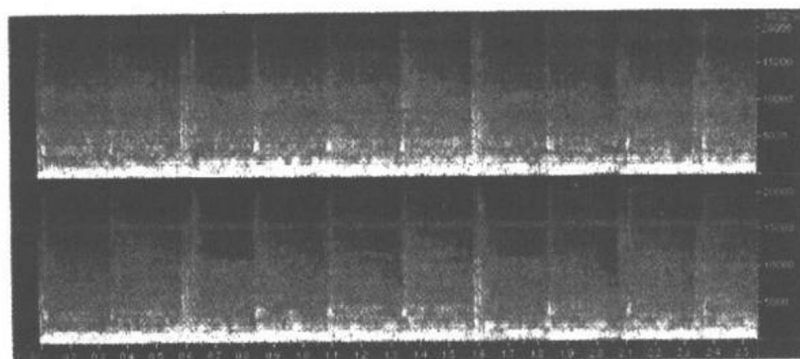


图1-4 一段音乐的原始频谱

前面讲过，MP3 在压缩时将裁剪掉那些人耳听不到，或不易听到的信号。图 1-5 就是图 1-4 中的 CD 音乐转换成 MP3 后的频谱。与图 1-4 相比较，可以很容易地看出，频率大于 15000Hz 的信号绝大部分已经被裁掉了。频率高于 15000Hz 的声音本来就不容易被人听到，何况是在频率较低的信号也存在的情况下。所以，我们很难察觉。

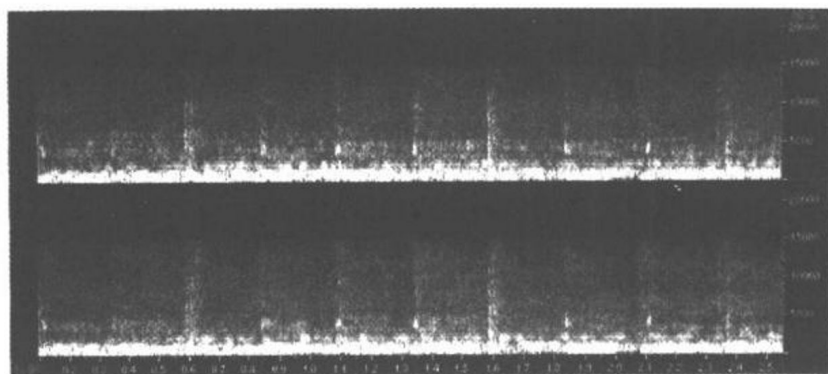


图1-5 音乐转换成 MP3 后的频谱

让我们再来看看上例中的音乐的原始声波图和转换成 MP3 后的声波图，分别如图 1-6 和图 1-7 所示。