



高成伟 编著
Gao Chengwei

Mobile Multimedia:
Standard, Theory and Applications

移动多媒体技术 ——标准、理论与实践



清华大学出版社

移动多媒体技术

——标准、理论与实践

高成伟 编著

清华大学出版社
北京

内 容 简 介

本书所讨论的具有实时要求的音视频多媒体应用将成为第三代移动网的主要数据业务。全书共分为五个部分,分别就音频编解码标准 MP3,语音编解码标准 AMR,合成音乐标准 MIDI,视频标准 H.267、H.264 和 MPEG-4,以及传输标准 3G-324M 展开详细的讨论。本书旨在解析这些移动多媒体标准中的技术难点,并与实际的应用相结合。书中不仅给出了多个空间实例,例如 MIDI 的音乐合成器及移动网上的双向视频通信系统等,而且对标准中所使用的信号处理技术提供了详细的理论证明。

本书既可用作通信、信号处理、信息、自动化、电子等理工科专业的高年级本科生和研究生的参考教材,又可以用作相关领域从业人员的参考书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

移动多媒体技术：标准、理论与实践/高成伟编著. —北京：清华大学出版社，2006.3

ISBN 7-302-12329-2

I. 移… II. 高… III. ①音频设备—多媒体技术—标准 ②图像通信—多媒体技术—标准
IV. ①TN912.2-65 ②TN919.85-65

中国版本图书馆 CIP 数据核字(2006)第 000810 号

出版者:清华大学出版社
http://www.tup.com.cn
社总机:010-62770175

地址:北京清华大学学研大厦
邮编:100084
客户服务:010-62776969

组稿编辑：王一玲

文稿编辑：田志明

印 装 者：三河市春园印刷有限公司

发 行 者：新华书店总店北京发行所

开 本: 175 × 245 印张: 16.75 字数: 353 千字

版次: 2006 年 3 月第 1 版 2006 年 3 月第 1 次印刷

书 号: ISBN 7-302-12329-2/TN · 298

印 数: 1 ~ 3000

定 价：28.00 元

作者简介

高成伟博士 1986 年毕业于北京清华大学自动化系,并于同年考入北京大学成为信息工程专业的硕士研究生。1989 年初赴美国华盛顿大学攻读博士学位,并获得电机工程及计算机工程的博士学位。

高成伟博士自 1992 年以来,长期从事音视频方面的编解码研究,并在多家公司主持音视频系统的设计和研发,在该领域发表多篇学术论文,取得多项成果和专利。高成伟博士服务于美国微软公司和美国高通公司等国际知名科技企业,并长期参与国际标准化组织(ISO)视频专家组(MPEG)的工作。

About the Author

Dr. Chengwei Gao was graduated, B. S. , from Automation Department of Tsinghua University, Beijing in 1986. In that year he was admitted into Beijing University as master degree student majoring in information technology. In the first half of 1989, Dr. Chengwei Gao was admitted into University of Washington, USA, as Ph. D. candidate, where he was granted Ph. D. in electrical engineering and computer engineering.

From 1992 onward, Dr. Chengwei Gao has focused on research area of video and audio codec. He has taken charge of design and implementation of video and audio systems in multiple corporations. On video and audio codec, Dr. Chengwei Gao has published many research papers. In the area of video and audio codec and system, Dr. Chengwei Gao has distinguished himself with a lot of achievements and has acquired multiple patents.

Dr. Chengwei Gao has served for companies in the United States, like Microsoft and Qualcomm. He has for a long time participated in the video standard work of MPEG.

前 言

Barnard 先生的一句名言“一像抵千言”(A picture is worth a thousand words.)形象地概括了信息变革给人类生活带来的巨大变化。随着 2.5G 移动网的逐渐完善和 3G 移动网的不断成熟,以及人们生活工作习惯的改变,移动多媒体应用必将成为移动数据业务的主角。人们可以随时随地享受娱乐,不受时间场所限制的工作环境。在移动通信中只闻其声,不见其人的时代即将成为过去。

移动数据业务在过去短短的十年间,已经经历了几代的变迁。2G 移动网的主要数据业务是短信服务(short message service, SMS),而 2.5G 由于移动手持设备处理功能的升级和网络带宽的增大,包含多媒体内容如声音、图片等的彩信业务(multimedia message service, MMS)成为了当今移动数据时代的标志应用。然而,以上这些数据业务都是非实时的数据业务。具有实时要求的多媒体业务,如视频电话、视频会议、音视频娱乐点播、移动实时监控等应用,将在 3G 移动时代实现。

工业标准的建立往往标志一个产业时代的到来,每项应用的适用范围及其对终端设备和运行环境的要求是不同的,有的甚至是很大的不同,如移动视频和高清晰度数字电视,它们都是视频应用,然而它们对终端设备和传输环境的要求却是大相径庭的。多媒体标准正在呈现出一种多样化趋势,且是以应用为导向的。本书的着重点是剖析移动多媒体的标准和核心技术。

多媒体信息是集文本、音频、视频、图形、图像、动画等信息为一体的综合表现形式。移动多媒体技术就是把上述媒体信息通过移动终端的数字化采样、压缩与存储,再经移动网络传输到另一个移动终端重现的过程。移动多媒体技术把多种信息集成在一起并通过移动网络提供交互和实时的媒体服务。移动多媒体应用的特点是,它所能支配的系统资源有限,如移动终端的 CPU 计算能力、存储或传输信息的带宽及电池的能量有限等。其中电池的能量是移动终端设备的“阿克里斯的后脚跟”(Achilles heel),是限制其发展的瓶颈。我们知道,多媒体信息的处理是十分复杂的,需要大量的计算,这就需要移动终端设备的 CPU 在高速下运行,可是 CPU 的运行速度与耗电量是成正比的,即 CPU 运行越快,电池的消耗也越大。我们很难想像不能播放一个完整节目的移动娱乐终端会赢得市场。当然,业界的领头羊企业也在不遗余力地设法解决这个问题,如 TI 的 OMAP 系列芯片、Qualcomm 的 6xxx 系列芯片及 Analog Device 的 Blackfin 系列芯片等均是优化的软硬件技术实现移动多媒体的复杂算法来达到减少功耗的目的,同时保证音视频的质量。然而,我们应该知道,所有的移动多媒体的解决方案都要基于对标准及其算法的深刻理解。本书将详

细分析移动多媒体的各种标准,帮助读者理解标准中各种核心技术的工作原理。本书的读者对象是通信、信号处理、信息、自动化、电子等理工科专业的高年级本科生和研究生及相关领域的从业人员。本书的重点是详细分析标准中的信息处理技术,因而读者应具有初等的信号处理知识。

全书共分为 5 章。第 1 章讨论移动多媒体的音频标准 MP3。MP3 包含了三个核心技术,即子带编码、修正的离散余弦变换(modified discrete cosine transform, MDCT)及心理声学模型,书中将重点对这三个技术进行分析并解释其工作原理。第 2 章介绍移动多媒体合成音乐标准——音乐数字接口协议(MIDI)。MIDI 在移动终端已得到很多应用,如手机铃声、游戏背景音乐及掌上卡拉 OK 等。MIDI 是把文本信息(text)转换成波形(waveform)音频样本的技术,本书将给出实现这种转换的 4 种数学模型:频率调制法(frequency modulation, FM)、限频法(subtractive synthesis)、波表法(wavetable)及波导法(waveguide)。第 3 章分析移动多媒体的语音编解码标准,自适应多码率(adaptive multi-rate, AMR)编解码技术。第 4 章讨论移动多媒体视频编解码标准 H. 263、H. 264 及 MPEG-4 技术,针对移动网误码率高及带宽有限的特性,详细讨论容错掩错技术及码率控制技术。第 5 章介绍移动多媒体的传输标准 3G-324M 并讨论双向视频通信在 3G 移动网上的应用。

本书如有不足之处,恳请读者批评指正。

高成伟

2005 年 6 月

于美国华盛顿州西雅图市

目 录

第 1 章 MP3 音频编解码技术标准	1
1.1 音频信号的采样及量化	1
1.2 音频信号的频域变换	5
1.3 信号的抽取与插值	8
1.4 两通道滤波器组	13
1.5 标准正交镜像滤波器组 QMF	15
1.6 共轭正交滤波器组 CQF	17
1.7 余弦调制滤波器组及伪 QMF	20
1.8 MP3 中的多通道滤波技术	21
1.9 MP3 PQMF 的实现	24
1.10 MP3 标准中的变换编码	29
1.11 修正的离散余弦变换	31
1.12 时频分辨率	39
1.13 DFT、PQMF 及 MDCT 之间的关系	43
1.14 MP3 标准中的心理声学模型	46
1.14.1 音频信号的掩蔽效应	46
1.14.2 听觉系统的临界频带	50
1.14.3 MP3 标准中的心理声学模型 I	51
1.14.4 MP3 标准中的心理声学模型 II	54
1.14.5 数据块类型的确定	60
1.14.6 基于心理声学模型的比特分配	62
1.15 MP3 的量化和熵编码	65
1.16 MP3 的立体声编码	68
1.17 MPEG-2 对 MP3 的兼容	69
1.18 小结	70
参考文献	71
第 2 章 MIDI 技术	74
2.1 频率调制音乐合成技术	76
2.1.1 频率调制法中音色的产生	78

2.1.2	音乐信号的包络线	79
2.1.3	频率调制法的音色实现	81
2.2	波表音乐合成技术	82
2.2.1	音色信号的周期分析与提取	82
2.2.2	波表法中的音调转换	84
2.2.3	音调转换的质量改善	87
2.2.4	使用波表法实现乐器音色	88
2.3	限频音色合成法	91
2.3.1	波形声源的谐波分析	91
2.3.2	限频法中的滤波系统	93
2.3.3	数字滤波器的实现	95
2.3.4	使用限频法实现音色	97
2.4	波导音色合成法	99
2.4.1	运动波的数学模型	99
2.4.2	使用波导音色合成法实现乐器音色	101
2.5	音响特效的合成技术	103
2.5.1	颤音效果的合成	103
2.5.2	回声音响特效的合成	104
2.5.3	多和声的合成	107
2.6	MIDI 标准	108
2.6.1	音符开始消息	109
2.6.2	音符终止消息	110
2.6.3	音色转换消息	110
2.6.4	音色特效消息	110
2.6.5	单音力度消息	111
2.6.6	信道力度消息	111
2.6.7	控制转换消息	111
2.7	小结	116
	参考文献	117
第3章	AMR 语音编解码标准	120
3.1	语音信号的预处理	123
3.2	线性预测分析	124
3.2.1	12.2 模式中的线性预测编码	128
3.2.2	其他模式中的线性预测编码	130

3.3	线性预测参数的线谱对变换	132
3.4	线性预测参数的线谱对量化	138
3.4.1	12.2 模式的线谱对量化	138
3.4.2	其他模式的线谱对量化	139
3.5	线性预测参数的线谱对的逆变换	140
3.6	线性预测参数的线谱对的插值	141
3.6.1	12.2 模式的线谱对插值	141
3.6.2	其他模式的线谱对插值	141
3.7	语音信号的听觉加权分析	142
3.8	语音信号的周期性分析与自适应编码	145
3.8.1	语音信号周期的开环搜索	146
3.8.2	语音信号周期的闭环搜索	147
3.9	语音信号的随机编码	154
3.10	小结	158
	参考文献	159
第4章	移动视频技术	161
4.1	视频信号基础	163
4.2	消除视频信号空域冗余的编码技术	168
4.2.1	多分辨率分析与小波变换	168
4.2.2	变换编码	179
4.2.3	基于帧内编码模式的能量预测及像素预测	183
4.3	消除视频信号时域冗余的编码技术	185
4.3.1	运动预测	186
4.3.2	运动补偿	191
4.3.3	预测运动矢量的编码及预测效率的提高	192
4.4	统计冗余的消除及熵编码	195
4.4.1	霍夫曼编码	197
4.4.2	算术编码技术	198
4.5	视频流码速控制	200
4.5.1	H.26x 中的码速控制	201
4.5.2	MPEG 中的码速控制	203
4.5.3	视频信号的转换编码与码速控制	206
4.6	视频编解码的容错及掩错	211
4.6.1	信道编码的容错	211

4.6.2 信源编解码的容错·····	212
4.6.3 解码端的掩错技术·····	219
4.7 小结·····	221
参考文献·····	222
 第5章 多媒体在3G移动网上的应用 ·····	227
5.1 流媒体在移动网上的传输·····	229
5.1.1 H.223 协议·····	230
5.1.2 H.223 协议复合层的容错技术·····	235
5.1.3 H.245 协议·····	238
5.1.4 多媒体逻辑信道·····	240
5.2 移动多媒体在移动网上的应用解析·····	241
5.3 小结·····	257
参考文献·····	258

第 1 章 MP3 音频编解码技术标准

音频信号是指人的听觉系统所能接收和辨别的通过空气传播的一种连续波,即声波。由于声波的压力与传播频率不同,所以人的听觉系统可以感知各种强弱不同、音调不同的声音。其实人的听觉系统即是一个带通滤波器,其通带范围在 $20\text{Hz} \sim 20\text{kHz}$ 之间;对于频率低于 20Hz 和高于 20kHz 的音频信号人的听觉系统便不能感知。我们知道,计算机不能处理连续时间域内的信号,为了使计算机能处理音频信号,必须对其进行采样,而音频标准正是针对离散的音频采样样本的处理技术与协议。工业界的音频标准主要包括 MPEG-1 Audio Layer 3(MP3), Advanced Audio Coding(AAC)与 Alias Cancellation Version 3(AC3)。AAC 与 AC3 是高保真、高采样率的音频标准,主要应用于高清晰数字电视、家庭影院及数字影院等领域,其应用背景决定了 AAC 与 AC3 对终端处理系统及传输网络系统都有较高的要求,因而 AAC 与 AC3 在移动多媒体领域内的商业应用还有待时日。尽管 AAC 也有低采样率版本,但本书不作讨论。

在所有音频标准中,MP3 的应用领域是最广阔的,如随身听、学习机、个人媒体播放器以及通信和广播等。MP3 已成为广为接受的通用音频编解码标准。在移动多媒体中,对 MP3 标准的实现已成为不可或缺的应用,MP3 包含了许多最新的对音频编码的研究成果,其中包括子带编码、MDCT 及心理声学模型等技术。在本章中,将详细解析 MP3 的核心技术。

1.1 音频信号的采样及量化

人的听觉系统所能感知信号的最高频率是 20kHz 。根据奈奎斯特采样定理可知,若采样频率 $f_s \geq 40\text{kHz}$,则我们可以不失真地重建音频信号;但理想重建音频信号的条件是原音频信号为带宽有限信号,且其最高频率 $f_{\max} \leq 20\text{kHz}$,否则在采样过程即会产生信号混叠,如图 1.1 所示。

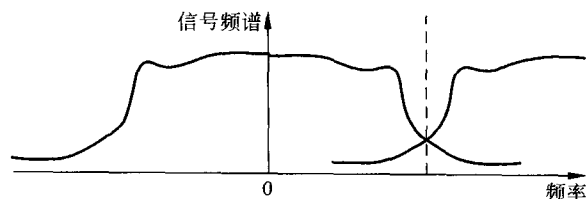


图 1.1 采样导致信号频谱混叠

为避免混叠发生,在对音频信号采样前,有必要对其进行滤波处理。我们可以用一 通带滤波器,其通带范围在 0Hz~20kHz 之间,这样以 40kHz 频率采样时,混叠将不会 发生,如图 1.2 所示。采样所得的样本通过使用脉冲编码调制(pulse code modulation, PCM)技术把模拟的连续信号转换为离散的数字信号,离散的数字信号通过 PCM 解码 器可还原成连续信号。图 1.3 和图 1.4 分别给出了模数及数模转换的过程。

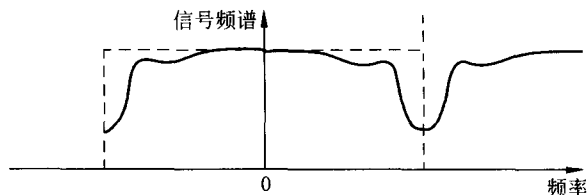


图 1.2 滤波处理可避免频谱混叠

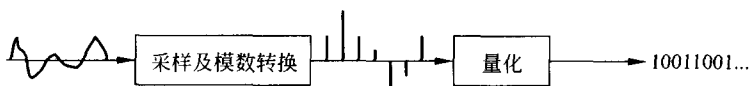


图 1.3 PCM 编码器

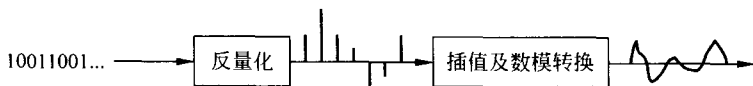


图 1.4 PCM 解码器

在得到音频样本后,要对其进行量化。在计算机的数字世界中,任何一个数值都是由 0 或 1 的排列组合,即二进制来表达的。所谓量化即是把采样样本的幅度信息用有限个[01]或比特组合来表示。由于模拟的幅度信号是用有限个比特来表示的,因此通过量化,无限多的模拟信号可映射为有限多的数字信号。由此可见,量化过程是有信号信息损失的。由于量化的技术、信号本身的动态范围及概率统计分布不同,因而由量化所引出的误差也不相同。

若用 R 个比特来编码音频信号的幅度样本,则可获得 2^R 个幅度级,即可把输入信号的幅度分布划分为 2^R 个子区间:

$$\zeta_k: \{x_k \leq x \leq x_{k+1}\}, \quad k = 0, 1, \dots, 2^R - 1$$

每个子区间由其标识 k 惟一确定。编码系统通过把子区间标识 k 变换为幅度值 y_k , 实现对样本的量化,如图 1.5 所示。图中, y_k 是重建信号值,而 x_k 则是判断阈值。这个量化过程可用量化函数来表示,即

$$y = Q(x) \quad (1.1)$$

其中 $Q(x)$ 为映射函数,而 $y_k = Q(x)$, $x \in \zeta_k$ 。

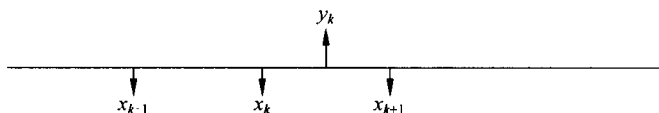
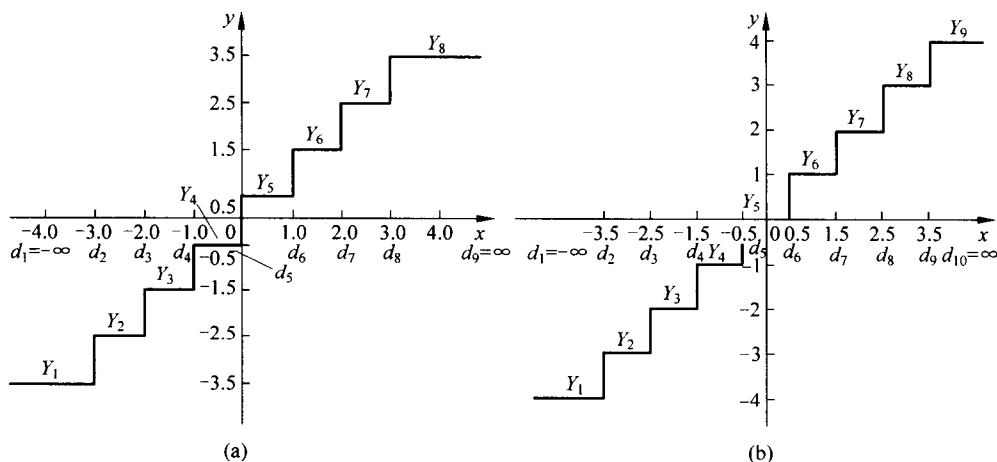


图 1.5 信号的量化

常用的 $Q(x)$ 函数有两种：不包含零值输出的 Midrise 映射(见图 1.6(a))和包含零值输出的 Midtreat 映射(见图 1.6(b))。

图 1.6 $Q(x)$ 函数

我们可以根据应用及信号特性来选择映射方法。我们知道,经采样后的音频信号的幅值有正有负,且其最大正负幅值相等。在量化音频信号时,我们一般选择对称量化,即在正负向上使用相同的比特数对幅值进行等量分级。Midrise 映射法不包含零值输出,它可把输入信号划分为偶数个子区间;而 Midtreat 映射法由于包含了零值输出,因此只能把输入信号划分为奇数个子区间。若用 R 个比特对输入信号进行量化,则最多可把输入信号划分为 2^R 个子区间。由于 2^R 是偶数,因此 Midrise 映射法可利用全部的 2^R 个子区间划分,而 Midtreat 映射法只能利用 $2^R - 1$ 个子区间划分。由于子区间与输出码字是一一对应的,因而 Midtreat 映射法较 Midrise 映射法损失一个码字。然而,Midtreat 映射法包含零值码字。在音频样本中,零值是重要的输出,没有零值,量化后的音频信号会出现明显的听觉噪声。

在输入信号的幅值取值范围的子区间划分上,可分为均匀量化与非均匀量化两种技术。均匀量化即指所划分的各子区间长度相等。不失一般性,假设输入信号 x 在 $-x_{\max}$ 与 $x_{\max}$ 间取值,即 $-x_{\max} \leq x \leq x_{\max}$,若用 R 个比特对其进行量化,则均匀量化技术把输入信号等分为 2^R (Midrise 映射)或 $2^R - 1$ (Midtreat 映射)个子区间,子区

间的长度依映射方法的不同而不同。Midrise 映射的子区间长度为 $2x_{\max}/2^R$ ；Midtreat 映射的子区间长度为 $2x_{\max}/(2^R-1)$ 。

前面已经指出,量化过程会引起信号损失,这种损失的大小用信噪比 SNR 来衡量,即

$$\text{SNR} = 10 \lg \frac{x_{\max}^2}{\sigma_q^2} \quad (1.2)$$

其中 σ_q^2 是量化误差的方差值。不失一般性,设定输入信号的取值范围为 $[-1, 1]$, 即 $x_{\max}=1$, 这样 SNR 可写成

$$\text{SNR} = -10 \lg \sigma_q^2 \quad (1.3)$$

由概率统计理论可知,量化误差 q 的方差 σ_q^2 由下式求得:

$$\sigma_q^2 = \int_{-\infty}^{\infty} q^2 p_q(q) dq \quad (1.4)$$

其中 $p_q(q)$ 为量化误差的概率密度分布。很显然,量化误差 q 满足 $|q| \leq \frac{\Delta}{2}$, 其中 Δ 为子区间长度。当子区间划分得足够细,即 Δ 够小时,可以认为量化误差 q 在 $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ 内均匀分布,因而

$$p_q(q) = \begin{cases} \frac{1}{\Delta}, & |q| \leq \frac{\Delta}{2} \\ 0, & \text{其他} \end{cases} \quad (1.5)$$

把式(1.5)代入式(1.4)中,可获得量化误差 q 的方差 σ_q^2 的解析形式为

$$\sigma_q^2 = \int_{-\infty}^{\infty} q^2 p_q(q) dq = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \frac{q^2}{\Delta} dq = \frac{\Delta^2}{12} \quad (1.6)$$

对 Midrise 映射, $\Delta = 2x_{\max}/2^R = 2^{-(R-1)}$; 对 Midtreat 映射, $\Delta = 2x_{\max}/(2^R-1) = 2/(2^R-1)$ 。当 R 足够大时, $2/(2^R-1) \approx 2^{-(R-1)}$, 因而无论采用哪种映射,都可认为量化的子区间长度为:

$$\Delta = 2^{-(R-1)} \quad (1.7)$$

把式(1.7)代入式(1.6),得到

$$\text{SNR} = -10 \lg \frac{2^{-2(R-1)}}{12} = 6.02R + 4.77 \quad (1.8)$$

由该式可看出,在量化端每增加一个比特,量化的信噪比则提高 6dB。

均匀量化是最简单的量化方式,其量化误差最大可达 $\frac{\Delta}{2}$, 因而当量化样本值在其可能取值空间呈均匀分布时,均匀量化能取得最佳效果。然而,在实际应用中情况并非如此,大多数信号的样本幅值都呈非均匀分布。在这种情况下,量化误差相对于取

值较小的信号样本的比例就会很大,从而使幅值较小的信号样本经量化后其重建精度变得很差,因此必须引入非均匀量化技术。非均匀量化技术对信号样本取值概率高的区域采用细量化,而对取值概率低的区域采用粗量化,因而,在所用量化比特数相等的条件下,非均匀量化比均匀量化能取得更小的均方差值。非均匀量化可以通过信号样本的非线性变换转化为均匀量化问题。记 x 为信号样本的幅值,且取值范围为 $[x_{\min}, x_{\max}]$, 可以找到一个非线性单调递增函数 $c(x)$, 使得 $y = c(x)$ 在 $[c(x_{\min}), c(x_{\max})]$ 区间内呈均匀分布。在上述非线性转换后,对原信号 x 的非均匀量化,即转换成对 y 的均匀量化(如图 1.7 所示)。在后面的章节中将会讲到,MP3 运用这种方法对音频频域信号进行非均匀量化。

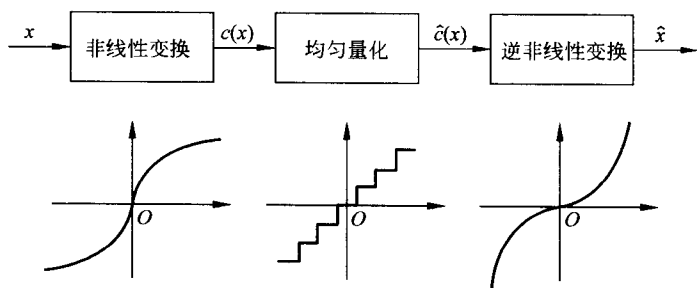


图 1.7 非均匀量化处理

1.2 音频信号的频域变换

在频域中比在时域中处理音频信号的好处要多。相邻音频信号间在时域内存在着较强的相关性,通过信号由时域到频域的转换,可以消除信号间的相关性,且频域信号依频谱由低到高,对人的听觉系统的影响从强到弱。频域内的音频信号具有能量的集中性(低频分量的能量较高频分量强)和影响力的有序性的特点,因而一般的音频处理都会转换到频域处理。MP3 音频处理的根本原则是建立在频域基础上的。下面介绍时域到频域的几种转换分析方法。

我们知道,任何一个时域内的连续周期信号,只要满足狄火赫利条件,皆可分解为傅里叶级数,且分解后的各分量的振荡频率是倍数关系,即谐波。各谐波的振幅与相位构成了周期信号的幅频特性和相频特性,分别称为幅度谱与相位谱。幅度谱与相位谱共同确定一个频率响应。对于时域内的连续周期函数 $x(t)$ (见图 1.8),其频率谱是离散的,且各频谱分量都是其基频的倍数,即有

$$x(t) = A_0 + \sum_{n=0}^{\infty} A_n \sin\left(\frac{n\pi t}{l} + \Phi_n\right) \quad (1.9)$$

其中, A_0 为直流分量, A_n 是 n 次谐波的振幅, Φ_n 是 n 次谐波的相位。 $x(t)$ 的周期为 $2l$, 基频为 $f=1/2l$ 。

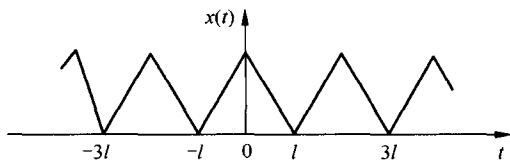


图 1.8 周期为 $2l$ 的周期函数

对于非周期连续时域信号 $x(t)$, 其频谱是连续而非离散的, 其傅里叶变换公式为

$$X(e^{j\omega}) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (1.10)$$

而其逆变换为

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) e^{j\omega t} d\omega \quad (1.11)$$

但不是所有时域信号 $x(t)$ 都存在傅里叶变换, 只有在时域内绝对可积的函数, 即 $\int_{-\infty}^{\infty} |x(t)| dt < \infty$, 才存在傅里叶变换。

由傅里叶变换可以看出, 时域信号 $x(t)$ 频谱的各分量皆是全局的, 即每个分量存在于信号的自始至终, 因此我们不知道什么时候哪个频率分量会出现, 为了弥补傅里叶变换的这一不足, 短时傅里叶变换被提出。短时傅里叶变换使用一个窗口函数把连续时间信号分割成相对独立的信号数据块, 这样就可以针对不同时间段信号的频谱分布进行分析。

在实际应用中, 我们所处理的信号形式是离散的而非连续的。在离散时间域内, 对应于连续时域内的傅里叶变换和短时傅里叶变换, 有离散时间傅里叶变换 (DTFT) 和离散傅里叶变换 (DFT)。

若离散时间信号 $x(n)$ 为能量有限信号, 即 $\sum_n |x(n)| < \infty$, 则其 DTFT 由下式给出:

$$X(e^{j\omega}) = \sum_n x(n) e^{-j\omega n} \quad (1.12)$$

由该式可以得知, 如果离散时间信号 $x(n)$ 为一实数序列, 那么其 DTFT $X(e^{j\omega})$ 是以 2π 为周期的周期函数, 任何 2π 区间都包含全部 $X(e^{j\omega})$ 的频谱信息。

$x(n)$ 的 DTFT 的逆变换可由下式求出:

$$x(n) = \oint_{2\pi} X(e^{j\omega}) e^{j\omega n} d\omega \quad (1.13)$$

DTFT 应用于整个时域信号序列,它对应于连续时间域的傅里叶变换。DFT 则是针对有限长离散信号序列的频域变换,对应于连续时间域的短时傅里叶变换。若有限离散时域信号 $x(n)$ 的长度为 N ,则其离散傅里叶变换由下式给出:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0, 1, \dots, N-1 \quad (1.14)$$

DFT 的逆变换由下式定义:

$$x(n) = \sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}nk}, \quad n = 0, 1, \dots, N-1 \quad (1.15)$$

DFT 变换在频域中所得的频率样本数与其时域样本数相等。

若用变量 z 替换 DTFT 中的 $e^{j\omega}$,则获得另一个在离散时间域内的重要变换工具 Z 变换,即有

$$X(z) = \sum_n x(n)z^{-n} \quad (1.16)$$

前面讲到,只有在时域内绝对可积的函数才存在傅里叶变换,对于不满足该条件的函数,如指数函数 e^{at} ($a > 0$),其傅里叶变换不存在。如果引入一个衰减因子 $e^{-\sigma t}$ ($\sigma > a$),那么 $x(t)e^{-\sigma t} = e^{-(\sigma-a)t}$ 是绝对可积的函数,因而 $x(t)e^{-\sigma t}$ 的傅里叶变换存在。令 $s = \sigma + j\omega$,并用 s 代替傅里叶变换中的 $j\omega$,得

$$X(s) = \int_0^{\infty} x(t)e^{-st} dt \quad (1.17)$$

这个变换称为拉普拉斯变换,其逆变换由下式定义:

$$x(t) = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} X(s)e^{st} ds \quad (1.18)$$

以上所介绍的这些变换方法都是在音频编解码 MP3 标准以及其他多媒体编解码标准中常用到的时频转换分析工具。这些变换都满足相似的特性,以傅里叶变换为例,本书中常用的变换特性如下。

(1) 线性特性

若

$$F\{x(t)\} = X(\omega), \quad F\{y(t)\} = Y(\omega)$$

则

$$F\{ax(t) + by(t)\} = aX(\omega) + bY(\omega)$$

(2) 时移特性

$$F\{x(t \pm t_0)\} = e^{\pm j\omega t_0} X(\omega)$$

(3) 频移特性

$$F\{x(t)e^{\pm j\omega_0 t}\} = X(\omega \mp \omega_0)$$

(4) 微分特性

$$F\left\{\frac{dx(t)}{dt}\right\} = j\omega X(\omega)$$