



高等学校电子与通信类专业“十二五”规划教材

语音与音频编码

主编 张雪英 贾海蓉
主审 张伟雄



西安电子科技大学出版社
<http://www.xduph.com>

高等学校电子与通信类专业“十二五”规划教材

语音与音频编码

主 编 张雪英 贾海蓉

主 审 张伟雄

西安电子科技大学出版社

内 容 简 介

本书系统地介绍了语音和音频编码的概念、原理、方法与应用。主要包括：语音与音频编码基础，语音信号数字模型及短时域分析，矢量量化，语音信号线性预测分析，语音编码，MPEG 音频压缩编码，环绕声编码标准及音频编码文件格式，家用音频设备中的纠错编码，常用的音频信号处理软件等。本书在阐明理论的基础上，介绍了音频编码技术在音频设备中的应用，达到了理论与实践的有机结合。

本书内容丰富，重点突出，原理阐述深入浅出，注重理论与实际应用的结合，可读性强。本书可以作为高等院校通信工程、电子工程、信息工程、计算机工程等专业高年级本科生相关课程的教材，也可供从事语音和音频编码技术研究的科研及工程人员参考。

图书在版编目(CIP)数据

语音与音频编码/张雪英，贾海蓉主编.

—西安：西安电子科技大学出版社，2011.2

高等学校电子与通信类专业“十二五”规划教材

ISBN 978-7-5606-2515-7

I. ① 语… II. ① 张… ② 贾… III. ① 语音数据处理—编码—高等学校—教材

IV. ① TN912.3

中国版本图书馆 CIP 数据核字(2010)第 257258 号

责任编辑 毛红兵 任倍萱

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467. 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 陕西华沐印刷科技有限责任公司

版 次 2011 年 2 月第 1 版 2011 年 2 月第 1 次印刷

开 本 787 毫米×1092 毫米 1/16 印张 15.75

字 数 371 千字

印 数 1~3000 册

定 价 23.00 元

ISBN 978-7-5606-2515-7/TP·1254

XDUP 2807001-1

*** 如有印装问题可调换 ***

本社图书封面为激光防伪覆膜，谨防盗版。

前言

语音与音频编码是一个跨学科、综合性的研究领域,涉及声学、心理学、语音学、数字信号处理、计算机技术与应用等学科。随着信息技术、多媒体技术的快速发展,以及业界日益增长的对语音与音频编解码应用人才的需求,高等学校里开设“语音与音频编码”课程的专业也越来越多,非常需要有相应的教材。本书就是为满足这样的需求而编写的。

本书内容体系完整,共分为九章。第一章为语音与音频编码基础,介绍语音与音频信号的数字化过程、心理声学模型以及感知编码原理,这些内容是本课程的基础。第二章为语音信号数字模型及短时时域分析,介绍语音的发音机理、数学模型,重点介绍语音信号的短时时域分析方法。第三章为语音信号的线性预测分析,主要介绍线性预测分析的原理以及方程组的解法及应用。第四章为矢量量化,介绍矢量量化的原理及相关的矢量量化器,重点介绍最佳矢量量化器的设计方法和降低复杂度的矢量量化器原理。第五章为语音编码,首先以语音编码的三种类型,即波形编码、参数编码和混合编码为主线,分别介绍每种编码类型最常用的、最流行的编码器原理;其次基于传统的窄带语音不能满足宽带信号在保持语音的自然度、听觉舒适性以及说话者在特定环境下现场感的不足,给出宽带变速率编码的必要性,并详细分析 AMR - WB 编码器的编解码原理。第六章为 MPEG 音频压缩编码,以感知音频编码为前提,详细介绍 MPEG 音频压缩编码的标准及主要的原理。第七章为环绕声编码标准及音频编码文件格式,主要介绍环绕声编码标准的两大阵营: Dolby(杜比)环绕声和 DTS 环绕声技术,分别介绍了具有代表性的 Dolby AC - 3、Dolby Digital Plus、Dolby TrueHD 以及 DTS 和 DTS HD 的关键技术及编解码方式;另外,还介绍了数字音频文件的格式,其中包括无损和有损压缩。第八章主要介绍家用音频设备中的纠错编码,首先,从目前音频设备当中存在纠错的事实,引出纠错编码的原理和相关定义,并简要介绍线性分组码、奇偶检验码和循环码的检错原理,为后面介绍 RS 码打下基础;其次,详细介绍了 RS 码的原理,并结合交叉交织技术,使得在家用音频设备中的纠错成为可能;最后,分别介绍了 CD、DVD、BD、HD DVD 的纠错编码原理。第九章为常用的音频信号处理软件,主要介绍 Cool Edit 和 Sound Forge 软件,并通过具体的实例对这两种软件进行阐述。

本书适合作为通信、电子信息类专业“语音与音频编码”(32 学时)的课程教材,也可供相关专业技术人员参考。为了便于读者更好地理解和掌握本书的内容及重点,本书在每章后均附有小结和习题,以帮助读者掌握要点和检验学习效果。

本书由张雪英教授和贾海蓉博士编写,张雪英统稿,张伟雄主审。具体分工是:第二、三、四、五章由张雪英编写,第一、六、七、八、九章由贾海蓉编写。

由于作者水平有限,疏漏和错误之处在所难免,敬请广大读者批评指正。

编者

2011 年 2 月

目 录

第一章 语音与音频编码基础	1
1.1 概述	1
1.2 语音与音频编码基础	1
1.2.1 声音信号的数字化	3
1.2.2 声音压缩编码的声学原理	8
1.3 小结	18
习题一	18
第二章 语音信号数字模型及短时域分析	20
2.1 概述	20
2.2 语音的发音机理	20
2.2.1 人的发音器官	20
2.2.2 语音生成	21
2.3 语音信号的数字模型	22
2.3.1 激励模型	22
2.3.2 声道模型	23
2.3.3 辐射模型	23
2.3.4 语音信号数字模型	24
2.4 短时域分析方法	25
2.4.1 语音信号的预加重处理	25
2.4.2 语音信号的加窗处理	26
2.4.3 短时平均能量	27
2.4.4 短时平均幅度函数	29
2.4.5 短时平均过零率	30
2.4.6 短时自相关函数	31
2.4.7 修正的短时自相关函数	34
2.4.8 短时平均幅度差函数	35
2.4.9 基于短时自相关法的基音周期估值	36
2.5 小结	37
习题二	37
第三章 语音信号线性预测分析	39
3.1 概述	39
3.2 语音信号线性预测分析的基本原理	39
3.3 线性预测误差滤波	40
3.4 模型增益 G 的计算	41
3.5 线性预测方程组的自相关解法	42
3.5.1 用 Levinson - Durbin 算法解自相关方程组	42
3.5.2 自相关法解线性方程组误差分析	43
3.6 LPC 导出的其他语音参数	43

3.6.1	LSP 的定义和特点	43
3.6.2	反射系数	44
3.6.3	对数面积比系数 LAR	45
3.7	小结	45
习题三		45
第四章	矢量量化	47
4.1	概述	47
4.2	矢量量化的基本原理	48
4.2.1	矢量量化的定义	48
4.2.2	失真测度	50
4.2.3	矢量量化器	50
4.3	最佳矢量量化器	51
4.4	矢量量化器的设计算法	52
4.4.1	LBG 算法	52
4.4.2	初始码书的选定与空胞腔的处理	55
4.5	降低复杂度的矢量量化系统	56
4.5.1	树形搜索矢量量化器	56
4.5.2	多级矢量量化器	58
4.5.3	波形/增益矢量量化器	59
4.6	小结	60
习题四		60
第五章	语音编码	61
5.1	概述	61
5.2	语音编码的分类及特性	61
5.2.1	波形编码	61
5.2.2	参数编码	62
5.2.3	混合编码	62
5.2.4	语音压缩编码的依据	62
5.3	语音编码技术的发展史	63
5.4	语音编码性能的评价指标	65
5.4.1	编码速率	65
5.4.2	编码质量	66
5.4.3	编解码延时	68
5.4.4	算法复杂度	68
5.5	语音信号波形编码	69
5.5.1	脉冲编码调制	69
5.5.2	自适应预测编码	74
5.5.3	自适应差分脉冲编码调制	76
5.6	语音信号参数编码	81
5.6.1	通道声码器	81
5.6.2	共振峰声码器	82
5.6.3	线性预测声码器	82
5.7	语音信号混合编码	88

5.7.1	合成分析技术和感觉加权滤波器	88
5.7.2	激励模型的改进	89
5.7.3	G.728 语音编码标准	90
5.8	语音信号宽带变速率编码	97
5.8.1	概述	97
5.8.2	AMR-WB 编码算法原理	99
5.8.3	AMR-WB 解码算法原理	118
5.9	小结	123
	习题五	123
第六章	MPEG 音频压缩编码	125
6.1	音频压缩编码的原理	125
6.1.1	音频压缩编码的必要性和可能性	125
6.1.2	感知音频编码原理	125
6.1.3	频域编/解码器原理	126
6.2	MPEG 音频压缩编码标准概述	130
6.3	MPEG-1 音频压缩编码的基本原理	134
6.4	MPEG-2 音频编码原理	148
6.5	MPEG-4 音频编码标准	154
6.6	小结	157
	习题六	158
第七章	环绕声编码标准及音频编码文件格式	159
7.1	概述	159
7.2	Dolby(杜比)环绕声编码标准	160
7.2.1	Dolby(杜比)AC-3 音频压缩算法	160
7.2.2	Dolby Digital Plus 编解码技术	167
7.2.3	Dolby TrueHD 编解码技术	169
7.3	DTS 环绕声编码标准	170
7.3.1	DTS 环绕声技术	170
7.3.2	DTS HD 环绕声技术	171
7.4	音频编码文件格式	172
7.4.1	概述	172
7.4.2	无损压缩的音频编码文件格式	173
7.4.3	有损压缩的音频编码文件格式	174
7.5	小结	179
	习题七	179
第八章	家用音频设备中的纠错编码	181
8.1	概述	181
8.2	家用音频设备中的纠错编码基础	182
8.2.1	纠错编码原理	182
8.2.2	奇偶校验码	183
8.2.3	线性分组码	183
8.2.4	循环码	186

8.2.5 交织处理	188
8.2.6 里德-所罗门编码	191
8.3 CD 的纠错编码	195
8.3.1 CD 盘的结构和数据记录原理	195
8.3.2 CD 盘中的交叉交织的里德-所罗门码	198
8.3.3 误码的隐蔽	199
8.3.4 CD 中的 EFM 调制	200
8.4 DVD 纠错编码和调制技术	202
8.4.1 DVD 中的纠错码	202
8.4.2 RSPC 和 RS 码的实验比较	204
8.5 蓝光 DVD	206
8.5.1 蓝光 DVD 的纠错码(Picket Code)	207
8.5.2 蓝光 DVD 的调制码(17PP 码)	210
8.5.3 蓝光 DVD 的其他调制码	212
8.6 HD DVD	213
8.6.1 HD DVD 的数据格式	215
8.6.2 HD DVD 的纠错编码	216
8.6.3 HD DVD 的调制(FSM)	217
8.7 小结	218
习题八	218
第九章 常用的音频信号处理软件	220
9.1 概述	220
9.2 Cool Edit 软件	222
9.2.1 Cool Edit 软件的特点	222
9.2.2 Cool Edit 软件的常用功能介绍	223
9.2.3 Cool Edit 插件介绍	225
9.2.4 用 Cool Edit Pro 软件制作音乐	230
9.3 Sound Forge 软件	241
9.3.1 Sound Forge 软件的特点及功能	241
9.3.2 Sound Forge 软件的缺点	241
9.4 小结	242
习题九	242
参考文献	243

第一章 语音与音频编码基础

1.1 概 述

随着科学技术的快速发展及生活水平的提高,人们对精神生活的要求也越来越高。人们通过电视机可以观看世界各地的奇闻轶事,通过影碟机可以欣赏扣人心弦的影片,通过便携式音乐播放器可以听到震撼人心的音乐。但是,在享受这一切的同时却会忽略了实现这一切的关键技术,那就是语音与音频信号的存储与还原,也即声音的录放技术。实际上,这一技术的存在与发展已经经历了百余年。1857年,法国发明家斯科特(Scott)发明了声波振波器,它是留声机的鼻祖。1877年,爱迪生发明了锡箔唱筒式留声机,记录了“玛丽抱着羊羔,羊羔的毛像雪一样白”,这总共8秒钟的声音成为世界录音史上的第一声,也正是从那时起音频技术开始出现。之后人们在不断地寻找新技术、新方法以便能够将声音信号更好地记录下来。从早期的钢丝钢带录音机、开盘式录音机,到后来的盒式磁带录音机,再到现在的激光唱片 CD(Compact Disk)、数字多用途盘 DVD(Digital Versatile Disc)、数字盒式磁带录音机 DCC(Digital Compact Cassette)、微型唱片 MD(Micro Disk)、MP3、MP4 以及最新的蓝光 DVD 等,各种录放设备的性能在不断增强,从而也极大地丰富了人们的生活。此外,语音与音频录放系统在电话留言、语音报警、会议录音、广播节目录制等众多领域都有着广泛应用。

早期的语音与音频系统是针对模拟声音信号的,它实现起来很容易,但是存在一些缺点。在语音与音频的录制、编辑和放音过程中容易混入各种噪声,使得原信号不能很好地恢复,从而导致声音信号的扭曲。随着信息技术的发展,数字信号处理在越来越多的领域得到了应用,数字声音信号应运而生。数字信号易于存储和远距离传输,没有累积失真,抗干扰能力强,数字化存储的信息可以被高品质地还原,这一系列的优点促使音频技术逐步向数字化方向发展,同样,语音与音频录放系统也在从模拟向数字声音系统发展。

1.2 语音与音频编码基础

语音与音频数字录放的基本过程为编码、存储与回放,这个过程中的一个非常关键的技术就是必须对声音信号进行编码,否则将需要相当大的存储容量或传输带宽来存储和传输信号,这样,数字技术的优点也就荡然无存了。例如一采样频率为 44.1 kHz、量化位数为 16 位的双声道数字音频信号,若不对其加以压缩,其传输速率达到 1.411 Mb/s,若用容量为 650 MB 的 CD-ROM 来存储,只能保存大约 7.68 分钟,约两首曲子。因此为了充分利用资源,就必须对信号进行压缩。

虽然语音编码与音频编码同属信源的压缩编码,但二者的编码算法是不同的。语音信号的来源只有一个,即人的发音器官,语音编码的基础是语音产生模型。然而,音频信号的来源包括了人耳能感觉到的所有声音,音源较多,信号复杂,无法用统一的声源模型来处理。尽管如此,各种音频信号是要被人耳接收的,因此可以利用人类听觉感知系统的特性来研究音频编码的方法,所以音频编码的基础是听觉模型。音频编码研究的问题是用最小的能量感知失真,以尽可能低的编码速率来表达音频信号,以便于音频信号的传输或存储。所谓用低速率对音频信号进行数字表达,就是设计一种音频编码或音频压缩算法,它能够在听觉意义上使信号失真降为最小,而不是仅仅使传统惯用的输入输出波形均方差为最小。音频压缩与音频编码具有相同的意义。

语音与音频编码的发展非常快,自从数字声音信号出现以来,便备受人们关注,特别是针对语音信号编码的研究发展较早也较为成熟。早在1972年,国际电电话咨询委员会(CCITT)就制定了64 kb/s脉冲编码调制(Pulse Code Modulation, PCM)语音编码的标准,即G.711标准,用于公用电话网。G.711标准建议采用A-律或 μ -律的瞬时压扩技术将8 kHz语音信号的13位或14位PCM码转换成8位二进制码。由于受当时编码技术和IC技术的限制,其后语音编码技术的发展停顿了较长时间。自从1984年公布32 kb/s自适应差分脉冲(Adaptive Differential Pulse Code Modulation, ADPCM)语音编码标准(G.721)以来,语音编码技术便得到了快速的发展,各种高效率编码标准不断涌现,其中包括G.723、G.728、G.729等。表1.1给出了几种主要语音编码标准的特点及应用领域。

表 1.1 几种主要语音编码标准的特点及应用领域

产生年份	描 述	比特率/(kb/s)	语音质量/MOS	应用领域
1972	G.711 PCM	64	4.4	公用电话网
1984	G.721 ADPCM	32	4.1	公用电话网
1996	G.723.1 MPLPC ACELP	5.3~6.3	3.9	数据通信
1992	G.728 LD-CELP	16	3.61	公用电话网
1995	G.729 CS-ACELP	8	3.92	公用电话网
1988	GSM RPE/LTP	13	3.7	数字移动通信
1989	CTIA VSELP	8	3.8	数字移动通信
1998	AMR-ACELP	4.75~12.2	与比特率有关	数字移动通信
2002	G.722.2 ACELP	6.6~23.85	与比特率有关	数字移动通信
2004	AMR-WB+ACELP 和 TCX	6~48	与比特率有关	数字移动通信

音频信号编码是先发展,后制定标准,虽然起步较晚,却得到了快速发展。从MPEG-1到MPEG-21,取得巨大成功的是MPEG-1音频,它是ISO\IEC批准的第一个高保真音频压缩标准,规定了3个不同层次的编码方案,其中层I多用于数字盒式磁带,层II常用于VCD和数字音频广播,层III则已成为网络音乐传输标准。随着人们对音乐认知水平的提高,传统的立体声已经不能满足要求,环绕立体声应运而生,比如Dolby AC-3、DTS、THX和MPEG-2等。这些技术不仅应用在家庭影院系统,也应用在影剧院,并且会在高

清晰度数字电视等系统中得到应用。表 1.2 给出了几种主要的音频编码标准。

表 1.2 几种主要的音频编码标准

产生年份	标准形式	输出比特率/(kb/s)	压缩比	应用领域
1992	MPEG-1 layer I	384	4 : 1	DCC
	MPEG-1 layer II	192	6 : 1~8 : 1	数字广播、CD-ROM 及 CD-I
	MPEG-1 layer III	64	10 : 1~12 : 1	ISDN 上的音频传输
1994	MPEG-2	64/每声道	11 : 1	DAB、DVD
1998	MPEG-4	2~64	据比特率计算	移动通信、Internet
1991	杜比 AC-3	32~640	据比特率计算	HDTV、DVD、家庭影院
2006	AVS	16~96	据比特率计算	移动通信、Internet

语音与音频编码不仅应用于数字声音录放系统，在数字声音通信系统中也占据着重要地位。另外，语音与音频编码技术的发展与成熟，为通信领域、广播电视领域和多媒体技术方面进一步研究编码技术的发展提供了理论依据。

1.2.1 声音信号的数字化

自然界的声是模拟信号，经过数字化处理后的声频信号必须还原为模拟信号，才最终转换成声音。其中，数字化处理过程包括采样、量化和编码，这一过程的处理直接影响到所恢复出来的信号的质量，是否能与原始的波形保持一致，是语音和音频数字化的基础，也是人们可以利用心理声学模型不影响听音效果的基础。数字技术使处理信息的能力大大提高，音频记录、信号处理和硬件设计的水平随着数字技术的发展而提高，有利于计算机的处理，更有利于向多媒体方向发展。

信息的数字化很早就进入了人们的生活，如明码电报中用数字化代码表示汉字就是一个实例。在音频信号当中，将时间域中幅度上连续变化的声音信号变换为脉冲数据的过程称为数字化。音频信号数字化的框图如图 1.1 所示。

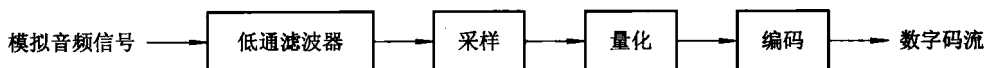


图 1.1 音频信号数字化框图

其中，低通滤波器的作用是去除高频频率分量，在数字音频系统的输出部分也要用低通滤波器来滤除系统内产生的高频成分，以重建原始的波形。

1. 采样

声波是声压幅度随时间连续变化的模拟量，它由传声器转换成声音信号后，是时间和电压幅度都连续变化的模拟信号。采样就是从时间上连续变化的声频信号中取出若干个有代表性的样本值，这些样本值能唯一地用来表征这一信号，并且能从这些样本中把信号完全恢复出来。采样的时间间隔称为采样周期，每秒内采样的次数称为采样频率，采样频率一般用 f_s 表示，声频信号的最高频率用 f_H 表示。当满足条件 $f_s \geq 2f_H$ 时，能够不失真地重建原模拟信号的波形，这就是著名的奈奎斯特(Nyquist)采样定理。 $2f_H$ 称为奈奎斯特频率。

1) 采样的数学描述

一个声频信号 $x(t)$, 经采样后的信号为 $x_s(t)$, ω 为原信号的角频率, ω_H 为信号最高角频率, ω_s 为采样角频率, T_s 为采样周期, 且 $\omega_s = 2\pi f_s = 2\pi/T_s$ 。

时域:

$$x_s(t) = x(t) \cdot \delta_T(t) \quad (1-1)$$

频域:

$$X_s(\omega) = \frac{1}{2\pi} [X(\omega) * \delta_T(\omega)] \quad (1-2)$$

有
$$X_s(\omega) = \frac{1}{T_s} \left[X(\omega) * \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_s) \right] = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_s) \quad (1-3)$$

采样的频谱图如图 1.2 所示。从图中可以看出: 原信号被采样之后的频谱 $X_s(\omega)$ 为原信号频率 $X(\omega)$ 按周期作重复延拓。不难看出, 只要满足 $f_s \geq 2f_H$, 在样值序列信号的频谱中可以完整地恢复出包含有原信号的频谱成分, 即包含原模拟信号的全部信息, 通过一个低通滤波器就能从样值信号中恢复出原信号。如果不满足奈奎斯特采样定理, 将会发生频谱混叠, 这样, 它就无法不失真地恢复原模拟信号, 由此产生的失真称为频谱混叠失真。如图 1.3 所示。

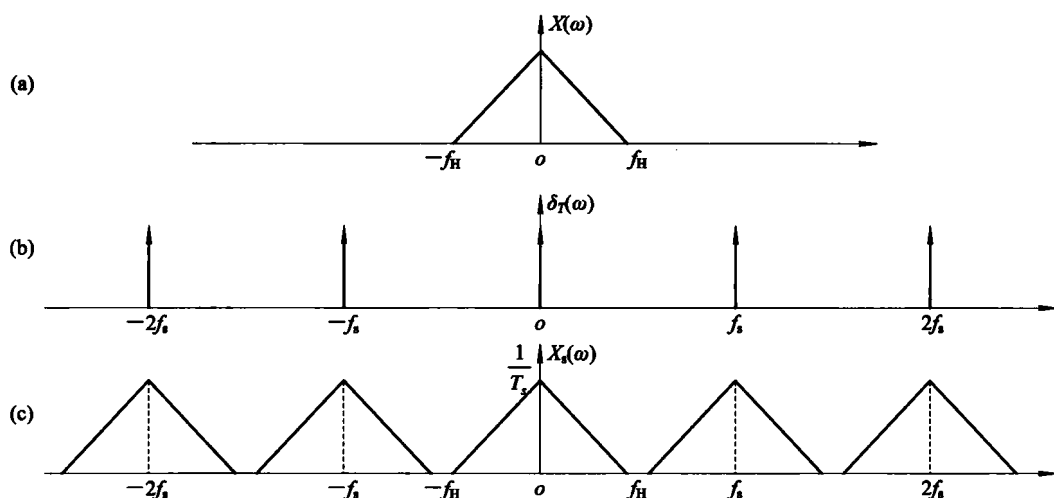


图 1.2 采样频谱变化图

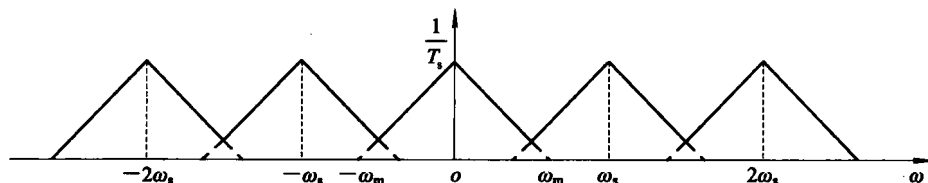


图 1.3 频谱混叠

2) 不失真恢复信号的条件

要避免信号发生频谱混叠失真, 就必须满足以下两个条件:

(1) 用低通滤波器限制带宽, 使 f_H 以上频谱分量为零。虽然人耳听不出 20 kHz 以上

频率的声音,但乐器频谱具有 20 kHz 以上的频率成分,若不去除,则会产生频谱混叠,所以要加一低通滤波器。

设采样频率 $f_s \geq 2f_H$, 且采样后使原信号变成理想脉冲串。即脉冲幅值与原信号一致,各脉冲宽度无限小。一般在音频信号中,是按照 $f_s = (2.1 \sim 2.5)f_H$ 来选取的。频率太高,会浪费带宽,降低有效性;频率太低,会发生混叠失真。目前,常用的音频采样频率有 32 kHz、44.1 kHz、48 kHz、96 kHz、192 kHz。采样频率与信号最高频率的关系不难做到,但理想脉冲串在现实中不存在,因为实际的脉冲都有一定的宽度。不过,当脉冲宽度为零时,采样输出也为零,所以现实的采样脉冲必须具有有限的时间宽度。另外,因为采样输出和脉冲宽度成比例,所以为了增大输出,确有必要采用一定时间宽度的脉冲。如果采样脉冲过宽,解调后的频率特性会在高频段下降,此即孔径效应。

(2) 插补用滤波器为理想滤波器,其截止频率为 f_b ,一般选取 $f_b = f_H$,通带内衰减为零,阻带的衰减为无穷大。实际上,数字系统无法达到上述理想状态,会产生一定误差。但是在设计音频设备时,应充分考虑实际滤波器与理想滤波器的差异,并进行适当补偿,才能使此种误差在听觉上不会造成问题。

2. 量化

采样是指把模拟信号变成时间上离散的脉冲信号,但脉冲信号的幅度仍然是模拟且连续的,因此还必须对其进行离散化处理,才能最终用数字信号来表示。这就要对幅值进行舍零取整的处理,即需要用有限个电平来表示模拟信号的抽样值,这个过程我们称之为量化。量化有两种形式:均匀量化和非均匀量化。下面将分别进行介绍。

1) 均匀量化

均匀量化是把输入信号的取值域按等距离分割。在均匀量化中,每个量化区间的量化电平均取在各区间的中间点,如图 1.4 所示。均匀量化的量化台阶 Δ 取决于输入信号的变

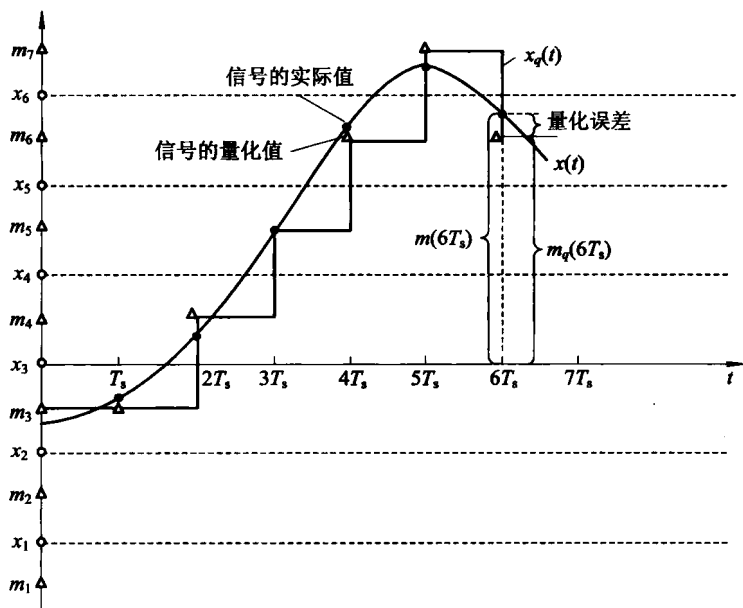


图 1.4 均匀量化原理图

化范围和量化电平数, 即当信号的变化范围和量化电平数确定后, 量化台阶也被确定。例如, 信号的最小值和最大值分别用 a 和 b 来表示; 量化电平数为 $M(M=2^n)$, 也称为量化级数, 其中 n 是量化比特数, 量化时采用二进制时的有效位数, 如 CD 中采用的是 16 bit, 它的量化级数就是 2^{16} , DVD 中采用的是 16 bit、20 bit 和 24 bit。那么, 其量化台阶为

$$\Delta = \frac{b-a}{2^n} \quad (1-4)$$

量化器的输出

$$x_q = m_i \quad (x_{i-1} < x \leq x_i) \quad (1-5)$$

式中, x_i ——第 i 个量化区间的终点, 可写成 $x_i = a + i\Delta$;

m_i ——第 i 个量化区间的量化电平, 可表示为 $m_i = \frac{x_i + x_{i-1}}{2} (i=1, 2, \dots, M)$ 。

信号功率与量化噪声功率之比是量化器的主要指标之一, 下面简单分析均匀量化时的信号量化噪声比。

在均匀量化时, 量化噪声功率 N_q 可由下式给出:

$$\begin{aligned} N_q &= E[(x - x_q)^2] = \int_a^b (x - x_q)^2 f(x) dx \\ &= \sum_{i=1}^M \int_{x_{i-1}}^{x_i} (x - m_i)^2 f(x) dx \end{aligned} \quad (1-6)$$

其中, $f(x)$ 为 x 的概率密度函数, 且

$$\begin{cases} x_i = a + i\Delta \\ m_i = a + i\Delta - \frac{\Delta}{2} \end{cases} \quad (1-7)$$

信号功率为

$$S = E[(x)^2] = \int_a^b x^2 f(x) dx \quad (1-8)$$

若已知随机变量 x 的概率密度函数, 便可计算出该值。

例如: 一个声音信号, 采用量化比特数 n 进行均匀量化, 其输入信号在 $[-a, a]$ 区间具有均匀概率密度函数, 试求该量化器平均信号功率与量化噪声功率之比。

由方程(1-6)得

$$\begin{aligned} N_q &= \sum_{i=1}^M \int_{x_{i-1}}^{x_i} (x - m_i)^2 \left(\frac{1}{2a}\right) dx \\ &= \sum_{i=1}^M \int_{-a+(i-1)\Delta}^{-a+i\Delta} \left(x - a - i\Delta + \frac{\Delta}{2}\right)^2 \left(\frac{1}{2a}\right) dx \\ &= \sum_{i=1}^M \left(\frac{1}{2a}\right) \left(\frac{\Delta^3}{12}\right) \end{aligned}$$

又因为 $M\Delta = 2a$, 所以

$$N_q = \frac{\Delta^2}{12}$$

又由式(1-8)得到信号功率

$$S = \int_{-a}^a x^2 \frac{1}{2a} dx = \frac{M^2 \Delta^2}{12}$$

因此, 信号量化噪声功率比为

$$\text{SNR} = \frac{S}{N_q} = M^2 \quad (1-9)$$

用 dB 表示时, 有

$$\text{SNR}(\text{dB}) = 20 \lg M = 20 \lg 2^n = 20n \lg 2 = 6.02n \quad (1-10)$$

其中, n 是量化比特即编码采用的比特位数。可以看出, 量化比特每增加一位, 信噪比提高约 6 dB。

信噪比的提高意味着声音动态范围的加宽, 若将量化比特数为 16 bit 的数字声记录在磁带上, 声音动态范围可扩展至约 98 dB, 接近于交响乐动态范围, 若将量化比特提高到 20 bit, 其声音动态范围可扩展至约 122 dB。由此看来, 量化比特数是个很重要的参数。要想提高信噪比, 提高量化比特数是一种可行的方法, 但是, 量化比特数不能无限制地提高, 它会带来量化级数的增加, 增大计算的复杂度。

2) 非均匀量化

非均匀量化是根据信号的不同区间来确定量化间隔的。对于信号取值小的区间, 其量化间隔也小; 反之, 量化间隔就大。与均匀量化相比, 它有两个主要的优点:

(1) 当输入量化器的信号具有非均匀分布的概率密度时, 非均匀量化器的输出端可以得到较高的平均信号量化噪声功率比, 并且, 非均匀量化时, 量化噪声功率的均方根值基本上与信号抽样值成比例。因此, 量化噪声对大、小信号的影响大致相同, 即改善了小信号时的量化信噪比。

(2) 非均匀量化一般用于声音信号, 这不仅因为其动态范围大, 也因为人耳在弱信号时对噪声很敏感, 在强信号时不宜觉察出噪声。

非均匀量化一般可分为 A-律和 μ -律量化。

3. 编码

采样、量化后的信号不是数字信号, 需要把它转换成数字脉冲编码, 这一过程称为编码。最简单的编码方式是二进制编码, 它用 n 比特二进制码来表示量化等级电平值, 每个二进制数对应一个量化电平, 然后把它们排列, 得到由二值脉冲串组成的数字信息流。用二进制数表示某一数值时, 该二进制数称为字。若以 8 位二进制数作为一个字, 则字内各位的名称如图 1.5 所示, 最左端的位称为最高有效位 MSB(Most Significant Bit), 以下依次是第二有效位(2SB), …… 第七有效位(7SB), 最右端的位称为最低有效位 LSB(Least Significant Bit)。

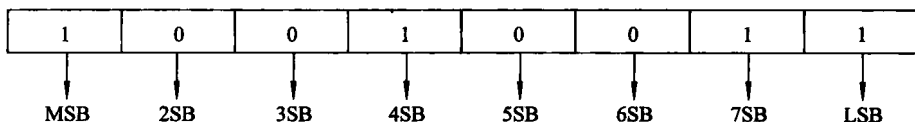


图 1.5 编码位数示意图

常用的二进制码主要有自然二进制码、偏移二进制码、补码、反射二进制码、格雷码等。

1.2.2 声音压缩编码的声学原理

1. 声音的特性参数

声音是通过空气传播的一种连续波,也可以说是机械振动或气流扰动引起周围弹性介质发生波动的现象,是声源(扬声器)振动引起的声波传播到听觉器官所产生的感受。所以说,声音是由声源振动、声波传播和听觉感受这三个环节所形成的。研究声音的目的是为了研究与声音有关的音频技术,研究音频技术是为了满足人们的听觉要求,而听觉不但取决于声音的特性,也与人的心理因素有关。由于人的个体差异,对声音这一客观现象的判断和感觉也有所不同,如对听觉的频率范围、对不同频率的感受程度以及对响度的反应等均有差异。所以,对声音进行定性分析是复杂的,而要对声音进行精确的定量分析更是相当困难的。下面先介绍一些与音频技术有关的声学参量。

1) 频率与倍频程

声波在每秒钟周期性振动的次数称为频率。频率与声音的对应关系是:频率低,相应的音调就低,声音就越低沉;频率高,相应的音调就高,声音就越尖锐。人耳可以听到的声音频率范围是 20 Hz~20 kHz,频率低于 20 Hz 的声音叫次声波,许多自然灾害如地震、火山爆发、龙卷风等在发生前都会发出次声波,次声波能够对人体造成危害,引起头痛、呕吐、呼吸困难等症状;高于 20 kHz 的声音叫超声波,医院中常用的 B 超检查就是把超声波射入人体,根据人体组织对超声波的传导和反射能力的变化来判断有无异常,如对人体脏器做病变检查、结石检查等,B 超检查具有对人体无损伤、简便迅速的优点。

倍频程定义为两个声音的频率或音调之比的对数,其公式为

$$n = \lg \frac{f_2}{f_1} \quad (1-11)$$

其中, f_1 为基准频率; f_2 为欲求其倍频程数的信号频率; n 为倍频程数,可正可负,也可以是分数或整数。例如, $n=1$ 称为“倍频程”, $n=1/2$ 称为“1/2 倍频程”。

两个频率相差 1 个倍频程,意味着其频率之比为 2^1 ;两个频率相差 2 个倍频程,意味着其频率之比为 2^2 ,……相差 n 个倍频程,意味着两个频率之比为 2^n 。按倍频程数均匀划分频率区间,相当于对频率按对数关系加以标度。

2) 声压与声压级

为了定量描述声音的强弱,引入两个概念,即声压和声压级。声压是由声波引起的交变压强,大气静止时存在一个压力,称为大气压。当有声波存在时,局部空间产生压缩或膨胀,在压缩的地方压力增加,在膨胀的地方压力减小。于是就在原来的静态气压上附加了一个压力的起伏变化。声压用 P 来表示,单位是 Pa(帕斯卡)。声压越大,声音就越大。但是人耳对声音强弱的感觉与声压的大小并不成线性关系,而是大体上与声压有效值的对数成正比,而且人耳能听到的声压范围在 $20 \mu\text{Pa} \sim 20 \text{ Pa}$ 之间,相差 10^6 倍,这给描述带来了不便,为了便捷描述且能适应人类听觉的特性,将有效声压与基准声压的比值取对数来表示声音的强弱,这种表示方式称为声压级,具体是指有效声压和基准声压比值取常用对数后的 20 倍,用 SPL 来表示,单位是分贝(dB, decibel),数学表达式为

$$\text{SPL} = 20 \lg \frac{P_e}{P_r} \quad (1-12)$$

其中, P_e 为有效值声压; P_r 为基准声压, 为 2×10^{-5} Pa, 是正常的青年人能听到的最弱声音。

3) 响度和响度级

无论是声压还是声压级, 都属于客观的物理量, 它们都是对客观事物的真实描述。另外, 对声音的描述, 除了客观方式外, 主观感受也是非常重要的, 描述声音强弱的主观感受的方式有响度和响度级。

响度表示人耳对声音大小及强弱的主观感受。响度主要依赖于引起听觉的声压, 但也与声音的频率和波形有关。响度的单位是“sone(宋)”, 国际上规定, 频率为 1 kHz、声压级为 40 dB 时的响度为 1 宋。1 宋 = 1000 毫宋, 1 毫宋约相当于人耳刚能听到的声音响度。

大量统计表明, 一般人耳对声压的变化感觉是声压级每增加 10 dB, 响度增加 1 倍, 响度与声压级有如下关系:

$$N = 2^{0.1(SPL-40)} \quad (1-13)$$

式中, N 为响度, SPL 为声压级。

表 1.3 列出了用这个公式计算出的部分响度与声压级的关系。

表 1.3 响度与声压级的关系

响度/宋	1	2	4	8	16	32	64	128	256
声压级/dB	40	50	60	70	80	90	100	110	120
响度级/方	40	50	60	70	80	90	100	110	120

声音响度级也是声音强弱的主观量, 即是凭人的听觉主观地判断声音强弱的量, 是人耳判断各种频率纯音响度级指标之一。响度级表示的是某响度与基准响度比值的对数值, 规定为等响度的 1 kHz 纯音(单一频率成分的音)的声压级, 单位是“方(phon)”。响度级为 40 方时, 响度为 1 宋, 响度级每增加 10 方, 响度增加 1 倍。

4) 音质

声音的质量由多种因素确定, 其中音调、音色、音量及音品是决定音频效果的四大要素。可以说音调由声波的频率决定, 音色由声波的频谱决定, 音量由声波的振幅决定, 而音品则由声波的波形包络决定。所有这些都是声音信号的物理量, 是可以进行客观技术测量的。

(1) 音调(Pitch)。音调表示声音频率的高低, 主要与声源每秒钟振动的次数有关, 是人耳对声调高低的主观评价尺度。它的客观评价尺度是声波的频率。音调低, 表示振动频率低, 声音显得深沉; 音调高, 表示振动频率高, 声音就尖刺。例如 C 的音符 6, 相当于 440 Hz, 而音符 6 $\flat$, 相当于 880 Hz, 音符 6 $\sharp$, 相当于 1760 Hz。

(2) 音色(Timbre)。音色是指声音的色彩和特点。不同的人 and 不同的乐器都会发出各具特色的声音, 可以说它与声源振动的频谱有关。如果说音调是单一频率的象征, 那么音色则是由多种频率所组成的复合频率的表现。图 1.6 所示为钢琴弹奏某一音阶时的声谱。由图可见, 这个声音的基频是 440 Hz, 除基频外, 至少包含有其他 15 种不同频率的振动。