

现代语音编码技术

吴家安 主编

科学出版社

北京

内 容 简 介

本书全面系统地阐述了现代语音编码的原理、技术及应用,讨论了目前语音编码的若干热点研究课题,介绍了语音编码的发展趋势和方向。全书由六个相对独立而又相互紧密联系的单元组成,共计十五章。主要内容包括现代语音编码技术导论、矢量量化编码、时域波形编码、子带编码、变换域编码、参数编码、混合激励线性预测编码、混合编码、多带激励编码、低速率和极低速率语音编码、变速率语音编码、宽频带高音质声频编码、感知音频编码、MPEG 音频压缩编码以及语音及音频编码的国际标准和地区标准等。本书取材新颖、结构合理、深入浅出、阐述清晰、内容丰富、实用性强,包含近二十几年来语音编码技术的许多新成果和新进展。

本书可作为通信、电子、计算机、信号处理、雷达、导航、无线电遥测、遥控、自动控制、广播、电视及其他信息技术专业的高等院校本科生及大专生的教学用书,也可供从事相关领域研究的科技工作者参考。

图书在版编目(CIP)数据

现代语音编码技术/吴家安主编.—北京:科学出版社,2007

ISBN 978-7-03-020507-0

I. 现… II. 吴… III. 语音数据处理—编码IV. TN912.3

中国版本图书馆 CIP 数据核字(2007)第 179171 号

责任编辑:刘宝莉/ 责任校对:陈玉凤

责任印制:刘士平/ 封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

印 刷

科学出版社发行 各地新华书店经销

*

2008 年 1 月第 一 版 开本:B5 (720× 1000)

2008 年 1 月第一次印刷 印张:19

印数:1—4 000 字数:363 000

定价:31.00 元

(如有印装质量问题,我社负责调换(环伟))

前 言

20 世纪 80 年代至今的二十几年间,是语音编码技术发展极为迅速的时期。现代语音编码技术主要是介绍在这二十几年间发展起来的新的语音编码技术,它现在仍然处于迅速发展之中,是通信和信息技术中备受关注的重要领域。现代语音编码技术的发展、进步和广泛应用,是现代通信技术领域中的一个闪光点,对于现代通信及信息技术的进步和发展起到了重要的推动作用。

众所周知,应用最早和最广泛的语音编码是 PCM (脉冲编码调制)。国际电报电话咨询委员会(CCITT)于 1972 年制定的 G. 711 建议规定的 A 律和 μ 律这两种 PCM 编码标准,都是非线性量化 PCM 编码方法。北美、日本使用 μ 律压扩方法,其他国家和地区使用 A 律压扩方法。编码器输出的速率通常为 64 kb/s 。PCM 是人们最熟悉的语音编码技术,而且 PCM 对于促进通信的数字化、推动通信和信息技术的迅速发展曾经起过极为重要的作用。但是,由于 PCM 的编码速率过高,占用信道带宽过大,不适应通信和信息技术发展的需要。为了压缩编码速率,减少传输占用的带宽,人们一直致力于研究开发新的语音编码技术。这种强烈的客观需求是推动语音编码技术发展的巨大动力。最近二十几年来,随着计算机、微电子、信号处理等相关技术的迅速发展和广泛应用,尤其是数字信号处理算法和器件(DSP 芯片)的飞速发展和应用,为中、低速率语音编码器的发展和应用准备了必要条件。正是在这种情况下,从 20 世纪 80 年代以来的二十几年间,语音编码技术进入一个飞速发展的时期。在这一时期,研究和开发出了 32 kb/s 、 24 kb/s 、 16 kb/s 、 8 kb/s 、 4.8 kb/s 、 4 kb/s 、 2.4 kb/s 、 2 kb/s 等速率的一系列语音质量良好的编码算法和编码技术,其中, 32 kb/s 、 16 kb/s 、 8 kb/s 、 4.8 kb/s 等速率的语音编码技术已经由 ITU-T (CCITT) 制定了国际标准,并进入了实用阶段;其余速率的语音编码技术也已经有了一些地区性标准,并正在成为制定新的国际标准的竞争热点,必将进一步推动语音编码技术沿着低速率和极低速率方向发展。

进入 21 世纪以来,语音编码技术开始进入一个新的全面、快速发展的时期。近年来在数字声广播、消费电子技术以及电话电视会议迅速发展的影响下,语音编码也在朝着宽频带高音质声频编码的方向发展,受到人们的广泛关注。其应用领域主要是数字声广播、电话电视会议系统、家电及娱乐电子技术。由于受到社会需求的强大推动,这是一个相当重要的发展动向。此外,第三代移动通信(3G)的语音编码技术是语音编码另一个备受关注的研究课题。3G 的三大主流技术 TD-SCDMA、WCDMA 和 CDMA 2000 都采用了变速率语音编码技术。TD-SCDMA 拟采用自适应

应多速率(AMR)语音编码;WCDMA 拟采用 AMR 语音编码,同时将可选模式声码器(SMV)作为备用技术;CDMA2000 选用变速率 CELP(QCELP)声码器或增强型变速率编码器(EVRC)。在数字卫星通信等系统中,也有的采用了变速率语音编码技术。变速率语音编码由于能够在保证重建信号质量的条件下进一步压缩编码速率,其应用前景十分广阔,是语音编码另一个重要的发展趋势。

近年来随着 DSP 芯片技术、语音学、语义学、神经生理学和心理学的迅速发展和广泛应用,发展起来一种利用人类听觉系统特性的感知编码技术,在音频压缩编码中得到广泛应用。感知编码器首先分析输入信号的频率和振幅,然后将其与人的听觉感知模型进行比较。感知编码器利用听觉感知模型去除那些与听觉感知无关的冗余部分,使得比特率大幅度降低,而编码信号质量并未下降。感知编码已经在 MPEG 音频压缩编码中应用,取得了良好效果,是语音编码又一个重要的发展动向。

国际标准化组织(ISO)和国际电工技术委员会(IEC)组织运动图像专家组(MPEG)于 1992 年 11 月制定了关于视频和音频信号压缩的国际标准 ISO/IEC 11172,即 MPEG-1 标准。这个标准有系统、视频和音频三大主要部分,最大音频比特率为 1.856 Mb/s。MPEG-1 中有关音频压缩的标准,即 11172-3,已经成功应用在 VCD、CD-ROM、ISDN、数字音频广播以及视频游戏等领域中,它支持每声道为 224~32 kb/s 的 32 kHz、44.1 kHz 和 48 kHz 的 PCM 数据。现在 MPEG-2、MPEG-4 等的音频层编码标准也已经制定出来,目前非常流行的 MP3、MP4 等都属于 MPEG 音频编码。因此,MPEG 音频编码也是一个十分值得注意的发展方向。

为了适应语音编码技术快速发展和应用的需要,从 20 世纪 80 年代至今的二十几年间,国际电报电话咨询委员会(CCITT)、国际电信联盟标准化部(ITU-T)和其他国际及地区标准化组织,制定了一系列的语音及音频编码的国际和地区标准。这些标准不但是学习、研究和应用现代语音及音频编码技术的重要依据,而且从另一个侧面反映出了这一时期语音及音频编码发展的历程和发展的趋势。

学习和掌握现代语音编码技术的基础知识、分析方法、关键技术和算法十分重要,只有了解和掌握这些已经成熟或基本成熟的方法和技术,才能很好地适应现代通信和信息技术发展的需要,做好我们的本职工作,也才能研究开发新的方案和算法,提出具有自主知识产权的技术方案。

本书编者根据多年从事语音编码教学和科研工作的经验、体会及成果,并参考了大量的文献和资料,试图总结现代语音编码技术的发展成就,介绍各种主要的实用编码技术及编码标准,全面、系统地阐述现代语音编码的原理、技术及应用,讨论目前的热点研究课题及最新成果,指出语音编码今后的发展趋势及方向,以期给广大读者以启示。

本书内容由相对独立而又相互紧密联系的六个单元共计十五章组成。第一单元是现代语音编码技术的入门知识,包括第一章“现代语音编码技术导论”和第二章“矢量量化编码”,分别介绍现代语音编码的基本知识、基本概念和基本技术。第二单元集中讨论经典的语音编码技术,即波形编码技术,包括第三章“时域波形编码”、第四章“子带编码”和第五章“变换域编码”,分别介绍时域波形编码、频域波形编码和变换域波形编码的原理、技术及应用。第三单元重点介绍现代参数编码技术,包括第六章“参数编码”和第七章“混合激励线性预测编码”,分别讨论 LPC 声码器基本原理、技术、应用及新的进展。第四单元较为详细地阐述语音混合编码的技术,包括第八章“混合编码”和第九章“多带激励编码”,分别介绍语音混合编码的基本原理、技术、应用及其新进展。第五单元讨论语音编码技术的几个热门研究课题,包括第十章“低速率和极低速率语音编码”、第十一章“变速率语音编码”、第十二章“宽频带高音质声频编码”、第十三章“感知音频编码”和第十四章“MPEG 音频压缩编码”,分别探讨了语音以及音频编码技术的几个主要发展趋势和方向。第六单元包括第十五章,介绍语音及音频编码的国际标准和地区标准,供读者在学习、研究和应用现代语音编码技术时参考。

本书可作为通信、电子、计算机、信号处理、雷达、导航、无线电遥测、遥控、自动控制、广播、电视及其他信息技术专业的高等院校本科生及大專生的教学用书,建议教学时数为 60 学时。书中打※号的章节内容,教师在教学过程中可以根据实际情况适当精简。如果使用本书对大專生教学时,可以省略这些章节的内容不讲。

本书由吴家安主编,吴家安、张会生、吴海锋、郭鸿基参编。吴家安、张会生制定了编写提纲和计划,吴家安编写了第一章、第三~十章和第十二~十五章的全部内容以及第二章、第十一章的部分内容,吴海锋编写了第二章的大部分内容,郭鸿基编写了第十一章的部分内容,吴海锋和郭鸿基绘制了书中的部分图表。全书由吴家安、张会生整理定稿。

本书在编写过程中,得到了西北工业大学的领导和老师的大力支持,电子信息工程专业的甄钊博同学绘制了书中的大部分图表,在此一并表示感谢。

由于编者水平有限,书中难免存在不妥之处,敬请读者不吝指正。

目 录

前言

第一章 现代语音编码技术导论	1
1.1 现代语音编码概述	1
1.1.1 语音编码和现代语音编码的概念及发展	1
1.1.2 对语音编码的要求	2
1.1.3 现代语音编码技术的作用和意义	2
1.1.4 语音编码系统的构成	3
1.2 语音编码的分类	4
1.2.1 传统的分类方法	4
1.2.2 按照编码速率的分类方法	5
1.2.3 按照被编码信号所在的域的分类方法	5
1.2.4 按照编码所使用的主要技术的分类方法	5
1.2.5 按照编码算法是否依赖于某种模型假定的分类方法	6
1.2.6 按照被编码信号的属性的分类方法	6
1.2.7 按照编码速率是否固定的分类方法	6
1.2.8 按照编码出现和应用时间以及技术发展情况的分类方法	7
1.3 衡量语音编码性能的主要指标	7
1.3.1 编码质量	7
1.3.2 编码速率	8
1.3.3 编译码复杂程度	9
1.3.4 编译码时延	9
1.3.5 坚韧性	10
1.4 提高语音编码质量的基本途径	10
1.5 语音编码发展现状及今后的发展方向	13
1.5.1 语音编码的发展现状	13
1.5.2 现代语音编码的发展趋势及方向	16
1.6 本书的内容结构安排	17
复习思考题	17
练习题	18
第二章 矢量量化编码	19

2.1	概述	19
2.1.1	量化的分类	19
2.1.2	矢量量化的发展及应用	19
※2.2	标量量化	21
2.2.1	无记忆标量量化	21
2.2.2	有记忆标量量化	34
2.3	矢量量化原理	37
2.3.1	矢量量化的定义	37
2.3.2	失真测度	39
2.3.3	矢量量化器的结构	41
2.3.4	矢量量化器的速率	41
2.3.5	最佳矢量量化器	42
2.3.6	矢量量化器的设计算法	44
2.4	无记忆矢量量化器	49
2.4.1	基本矢量量化器	50
2.4.2	树搜索矢量量化器	52
2.4.3	多级矢量量化器	59
2.4.4	乘积码矢量量化器	64
※2.5	有记忆矢量量化器	68
2.5.1	反馈矢量量化器	68
2.5.2	自适应矢量量化器	70
	复习思考题	71
	练习题	71
第三章	时域波形编码	73
3.1	PCM	73
3.1.1	PCM 的基本原理	73
3.1.2	PCM 时分复用原理	78
3.1.3	PCM 的 A 律 13 折线编码	82
3.2	DPCM	88
3.3	ADPCM	90
3.4	DM 和 ADM	91
3.4.1	DM	91
3.4.2	ADM	92
3.5	CCITT G. 721 建议的高质量 32 kb/s ADPCM	92
3.6	CCITT G. 727 建议的镶嵌式 ADPCM	99

复习思考题	101
练习题	102
第四章 子带编码	103
4.1 概述	103
4.2 子带编码的工作原理	104
4.3 子带编码的比特分配	106
4.4 整数带滤波器组	109
※4.5 正交镜像滤波器组	111
4.6 7kHz 带宽 SB-ADPCM 高音质声频编码系统	112
4.6.1 CCITT G. 722 关于 7 kHz 带宽高音质声频编码方案的主要内容	112
4.6.2 编、译码器的组成	113
4.6.3 CCITT G. 722 编码方案的特点	116
4.6.4 G. 722 编码标准的应用场合	116
复习思考题	117
练习题	117
第五章 变换域编码	118
5.1 变换编码概述	118
5.1.1 什么是变换编码? 为什么进行变换?	118
5.1.2 正交变换与正交矩阵	118
5.2 几种常用的正交变换	120
5.2.1 K-L 变换	120
5.2.2 W-H 变换	122
5.2.3 DCT (离散余弦变换)	124
5.2.4 Haar 变换	127
5.3 变换域系数的量化和编码	129
5.3.1 变换矩阵的选择	129
5.3.2 量化系数的选择	130
复习思考题	130
练习题	131
第六章 参数编码	132
6.1 参数编码概述	132
6.1.1 语音信号的产生模型	132
6.1.2 参数编码的类型	134
6.2 线性预测原理	136
6.2.1 预测编码的基本原理	136

6.2.2	线性预测	138
6.2.3	线性预测方程组及预测系数的确定	139
※6.3	线性预测的几种推演参数	140
6.3.1	反射系数	141
6.3.2	对数面积比系数	141
6.3.3	倒谱系数	141
6.3.4	预测器多项式的根	142
6.3.5	全极点系统的冲激响应	142
6.3.6	全极点系统冲激响应的自相关系数	143
6.3.7	预测误差滤波器冲激响应的自相关系数	143
※6.4	线谱对	143
6.4.1	线谱对分析的基本原理	143
6.4.2	线谱对参数的求解	145
6.5	线性预测声码器	145
6.5.1	考虑语音短时和长时相关性的语音生成模型	146
6.5.2	LPC 声码器	147
6.5.3	特征参数的提取	148
6.6	LPC-10 声码器	150
6.6.1	LPC-10 的编码器	150
6.6.2	LPC-10 的译码器	151
6.6.3	LPC-10 参数的编码与译码	152
6.7	LPC-10e 声码器	153
6.7.1	LPC-10e 的编码器	153
6.7.2	LPC-10e 的译码器	155
6.8	LPC 声码器存在问题及其改进措施	156
6.8.1	LPC 声码器存在的主要问题	156
6.8.2	LPC 声码器的改进措施	157
	复习思考题	158
	练习题	159
第七章	混合激励线性预测编码	160
7.1	概述	160
7.2	MELP 声码器编码原理	160
7.3	MELP 声码器译码原理	166
7.3.1	基音周期的译码	166
7.3.2	增益的译码和抑制	166

7.3.3	参数的插值	167
7.3.4	混合激励的生成	167
7.3.5	自适应谱增强	167
7.3.6	线性预测合成	168
7.3.7	增益校正	168
7.3.8	脉冲整形滤波	168
	复习思考题	168
	练习题	169
第八章	混合编码	170
8.1	语音混合编码的一般原理	170
8.1.1	感觉加权滤波器	170
8.1.2	合成分析法	172
8.1.3	激励源的改进	173
8.2	多脉冲激励线性预测编码	173
8.2.1	多脉冲激励线性预测编码的原理框图	173
8.2.2	最佳激励脉冲参数的估值	174
8.2.3	准最佳激励脉冲参数的估值	175
8.2.4	多脉冲线性预测编码的应用	176
8.3	规则脉冲激励线性预测编码	176
8.3.1	RPELPC 的基本原理	176
8.3.2	RPE-LTP 的编码方案	178
8.3.3	GSM 系统的 RPE-LTP 编码器	179
8.3.4	GSM 系统的 RPE-LTP 译码器	180
8.4	码激励线性预测编码	180
8.4.1	CELPC 的基本原理	181
8.4.2	美国政府标准(FED-STD-1016)4.8 kb/s 的 CELPC 声码器	182
8.4.3	美国电子工业协会电信协会(EIA/TIA)标准 8 kb/s 的 VSELPC 声码器	184
8.4.4	CCITT G. 728 建议的 16 kb/s 低延时 CELPC(LD-CELPC)声码器	186
8.4.5	ITU-T G. 729 建议的 8 kb/s 共轭结构代数码激励线性预测(CS-ACELP)声码器	189
	复习思考题	191
	练习题	191
第九章	多带激励编码	192
9.1	概述	192

9.2	MBE 语音模型	193
※9.3	MBE 语音参数的估计	197
9.3.1	某个子频带内的 MBE 语音产生模型	197
9.3.2	MBE 模型参数的估计	198
9.3.3	MBE 语音合成	199
9.4	MBE 声码器的应用	200
	复习思考题	200
	练习题	201
※第十章	低速率和极低速率语音编码	202
10.1	低速率语音编码概述	202
10.2	波形内插语音编码	203
10.2.1	狭义波形内插法和广义波形内插法	203
10.2.2	浊音语音的波形内插	203
10.2.3	广义内插(GWI)模型	207
10.2.4	实用 GWI 语音编码系统	209
10.3	正弦变换编码	213
10.3.1	语音信号的正弦分析	213
10.3.2	语音信号的正弦合成	214
10.3.3	低速率正弦编码器	215
10.3.4	与原型波形内插相结合	216
10.4	极低速率语音编码概述	217
10.5	1200~ 400 b/s 的语音编码技术	218
10.5.1	帧填充技术	218
10.5.2	利用矢量量化技术	218
10.6	400 b/s 以下的语音编码技术——语音识别与合成技术	219
10.6.1	识别合成型声码器的基本原理	219
10.6.2	关于识别合成型声码器编码速率的估计	220
	复习思考题	220
※第十一章	变速率语音编码	222
11.1	概述	222
11.1.1	变速率语音编码的必要性和可能性	222
11.1.2	变速率语音编码中使用的相关新技术	223
11.1.3	变速率语音编码的速率控制方式	223
11.2	实现变速率语音编码的关键技术	224
11.2.1	语音激活检测(VAD)技术	224

11.2.2	速率判决(RDA)技术	225
11.2.3	差错隐藏(ECU)技术	225
11.2.4	舒适背景噪声(CNA)生成技术	225
11.3	增强型变速率语音编码	226
11.4	变速率 CELPC(QCELPC)语音编码	227
11.4.1	QCELPC 概述	227
11.4.2	QCELPC 编码器	228
11.4.3	QCELPC 译码器	232
11.5	自适应多速率(AMR)语音编码	233
11.6	可选模式声码器(SMV)	235
	复习思考题	236
	练习题	236
※第十二章	宽频带高音质声频编码	237
12.1	宽频带声频的发展状况	237
12.2	宽带声频编码技术	238
12.2.1	宽带声频波形编码的分类	238
12.2.2	CCIR 建议的宽频带声频编码方式	239
12.3	宽频高音质声频编码器实例	239
12.3.1	编码系统组成	239
12.3.2	设计特点	240
12.3.3	编码系统的工作原理	240
	复习思考题	241
※第十三章	感知音频编码	242
13.1	心理声学基础	242
13.1.1	人类听觉系统的听觉阈值	242
13.1.2	频域掩蔽	242
13.2	感知音频编码/译码系统的组成	245
13.3	信宿(听觉)模型	245
13.3.1	单音的听觉模型	245
13.3.2	考虑掩蔽效应的听觉模型	247
13.4	感知音频编码的时-频分析信源模型	248
13.5	根据听觉(信宿)模型进行量化编码	248
13.6	感知音频编码器	248
13.6.1	音频编码-1(AC-1)	248
13.6.2	音频编码-2(AC-2)	249

13.6.3	音频编码-3(AC-3)——数字杜比	249
13.6.4	Apt-X100 音频编码器	254
	复习思考题	254
第十四章	MPEG 音频压缩编码	255
14.1	MPEG 音频编码概述	255
14.1.1	声音	255
14.1.2	模拟音频	256
14.1.3	数字音频	256
14.2	MPEG-1 音频压缩编码	256
14.2.1	MPEG-1 音频层概述	256
14.2.2	MPEG-1 音频层的数据结构	257
14.2.3	心理声学模型	257
14.2.4	MPEG-1 的音频层第 I 层	260
14.2.5	MPEG-1 的音频层第 II 层	261
14.2.6	MPEG-1 的音频层第 III 层	263
14.2.7	MP3	265
14.3	MPEG-2 音频压缩编码	267
14.3.1	MPEG-2 音频压缩编码概述	267
14.3.2	MPEG-2 音频编码与译码	267
14.3.3	MPEG-2 AAC(高级音频编码)	269
14.4	MPEG-4 音频压缩编码	271
14.4.1	MPEG-4 音频压缩概述	271
14.4.2	交互操作	271
14.4.3	音频编码	272
	复习思考题	274
	练习题	274
第十五章	语音及音频编码的国际标准和地区标准	275
15.1	概述	275
15.1.1	主要的国际标准化组织	275
15.1.2	主要的地区标准化组织	277
15.2	各标准化实体制定语音编码标准的一般工作程序	278
15.2.1	建立参考条款	278
15.2.2	语音编码器的性能测量	279
15.2.3	工作日程	280
15.3	CCITT(ITU)已经公布的电话带宽语音编码标准	280

15.4	ITU-T 近十几年来公布的电话带宽语音编码标准	281
15.5	CCITT 已经公布及 ITU-T 计划的宽带语音编码标准	281
15.6	国际海事卫星组织 4.15 kb/s 的 IMBE 语音编码标准	282
15.7	各主要地区标准化组织的语音编码标准	282
15.7.1	北美数字蜂窝移动通信语音编码标准	282
15.7.2	欧洲数字蜂窝移动通信语音编码标准	283
15.7.3	美国的保密通信语音编码标准	283
15.7.4	日本的数字蜂窝移动通信语音编码标准	284
15.8	MPEG 音频压缩编码标准	284
15.8.1	MPEG-1 音频压缩编码标准	284
15.8.2	MPEG-2 音频压缩编码标准	285
15.8.3	MPEG-4 音频压缩编码	286
	复习思考题	287
	参考文献	288

第一章 现代语音编码技术导论

最近二十几年来,随着数字通信、计算机、信号处理、微电子等相关技术的发展和广泛应用,语音编码技术发展非常迅速,取得了一系列突破性的成果,极大地促进了数字通信的发展和普及,是现代通信以及信息技术的一个亮点。所谓现代语音编码技术,主要就是指最近二十几年发展起来并得到广泛应用的语音编码技术。本章主要介绍现代语音编码技术的基本概念和基本知识,如什么是语音编码、对语音编码的要求、现代语音编码技术的作用和意义、现代语音编码系统的构成、分类以及主要性能指标等,以使读者对现代语音编码有一个初步的概要的了解。这些基本概念和基本知识,是学习和研究现代语音编码技术的入门知识和重要基础。读者应深刻理解,牢固掌握。

1.1 现代语音编码概述

1.1.1 语音编码和现代语音编码的概念及发展

在现代通信中,随着科学技术的迅速发展,图像、数据等非语音信息在通信信息总量中所占的比例大大提高,而且这种提高的趋势仍然会继续下去。但是,到目前为止,在大多数通信系统中,传输最多的信息仍然是语音信号。在可以预见的未来的通信中,尽管语音信号在通信信息总量中所占的比例会有所下降,但仍然会是传输最多的信息。

语音信号是模拟信号,不能直接在数字通信系统中传输,必须首先进行模/数转换和数/模转换。这种模/数转换和数/模转换就称为语音编译码,其作用是在发送端将语音模拟信号转换为数字信号,到了接收端,再将收到的语音数字信号还原为语音模拟信号。语音编译码又简称为语音编码。可见,语音编码技术在数字通信中具有十分重要的关键的作用。

语音编码属于信源编码。自从20世纪70年代以来的三十多年中,尤其是20世纪80年代至今的二十几年来,随着科学技术的迅速发展,特别是随着计算机技术、微电子技术、信号处理技术、语音学、语义学、神经生理学和心理学以及编码理论的发展和进步,语音编码技术取得许多突破性进展,研究出许多实用的编码方案,这些方案在不断研究、改进和应用中日趋成熟,形成了各种实用的语音编码技术。现代语音编码技术已成为通信技术中一个相当重要的学科,在各种通信网络中

都得到广泛应用。

现代语音编码技术主要是指从 20 世纪 80 年代至今二十几年间发展起来的新的语音编码技术,它现在仍然处于迅速发展之中,是通信和信息技术中备受关注的重要领域。

1.1.2 对语音编码的要求

各种通信网络工作的环境各不相同,传输信息也不完全一样,用户情况千差万别,因而对语音编码提出的要求也就不尽相同。综合各种通信网络对语音编码的共同要求,大致有以下几点:

(1) 编码速率要适合在常用话音信道内传输,一般要求编码速率在 16~ 2 kb/s 范围内的情况较多。

(2) 在一定编码速率下,语音质量应尽可能高,即译码后恢复语音的保真度要尽量高,一般要求达到长话质量(网络质量),即 MOS 评分 4 分以上,或要求达到通信质量,即 MOS 评分 3.5 分左右。

(3) 编译码时延要小。总时延一般要求不大于 65 ms。

(4) 编译码算法复杂度不能太大,以适于用大规模集成电路实现。

(5) 坚韧性好,有较好的抗误码性能。

上述这些要求之间往往是互相矛盾的。例如,为了使语音质量好,编码速率就应高一些,但这又会使其占用的信道带宽增大。信道带宽是有限的,编码速率过高就不能在信道内传输。因此,在实际应用中要根据具体情况综合分析和比较,选择最佳的编码方案。

1.1.3 现代语音编码技术的作用和意义

应用最早和最广泛的语音编码是 PCM(脉冲编码调制)。CCITT 于 1972 年制定的 G. 711 建议规定的 A 律和 μ 律这两种 PCM 编码标准,都是非线性量化 PCM 编码方法。北美、日本使用 μ 律压扩方法,其他国家和地区使用 A 律压扩方法。其编码器输出的速率通常为 64 kb/s。由于 PCM 的编码速率过高,不适应通信和信息技术发展的需要,为了压缩编码速率,减少传输占用的带宽,人们一直在致力于研究开发新的语音编码技术。这种强烈的客观需求是推动语音编码技术发展的巨大动力。另一方面,最近二十几年来,随着计算机、微电子、信号处理等相关技术的迅速发展和广泛应用,尤其是随着数字信号处理算法和器件(DSP 芯片)的飞速发展和应用,为中、低速率语音编码器的发展和应用准备了必要条件。正是在这种情况下,从 20 世纪 80 年代以来的二十几年间,语音编码技术进入一个飞速发展的时期。在这一时期,研究和开发出了 32 kb/s、24 kb/s、16 kb/s、8 kb/s、4.8 kb/s、4 kb/s、2.4 kb/s、2 kb/s 等速率的一系列语音质量良好的编码算法和编码技术,

其中,32 kb/s、16 kb/s、8 kb/s、4.8 kb/s 等速率的语音编码技术已经由 ITU-T (CCITT)制定了国际标准,并进入了实用阶段;其余速率的语音编码技术也已经有了一些地区性标准,并正在成为制定新的国际标准的竞争热点,必将进一步推动语音编码技术的发展。

现代语音编码技术就是指 20 世纪 80 年代以来发展起来的这一系列新的语音编码技术。这些新的语音编码技术的出现,极大地推动了通信和信息技术的发展,是现代通信发展史中的一个闪光点。

学习和掌握现代语音编码技术的基础知识、分析方法、关键技术和算法十分重要,只有了解和掌握这些已经成熟或基本成熟的方法和技术,才能很好地适应现代通信和信息技术发展的需要,作好我们的本职工作,也才能研究开发新的方案和算法,提出具有自主知识产权的技术方案。

1.1.4 语音编码系统的构成

目前语音编码已获得非常广泛的应用,语音编码系统的构成多种多样。但是,归纳起来可以分为两大类:第一类,编码-存储-回放系统,又称为数字语音录放系统;第二类,编码-传输-译码系统,又称为数字电话通信系统。图 1.1.1(a)和(b)分别示出了这两种语音编码系统的原理框图。

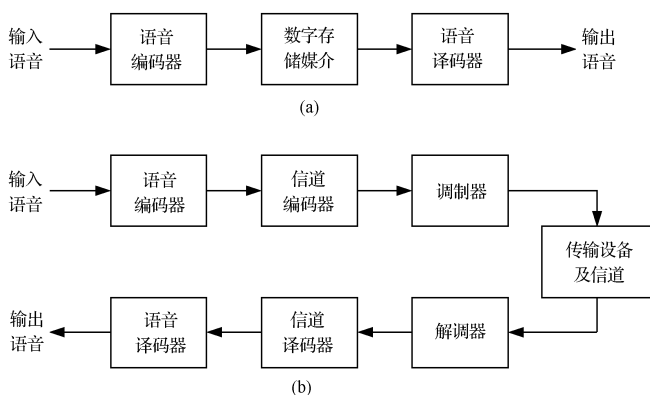


图 1.1.1 两种语音编码系统原理框图

数字语音录放系统与模拟语音录放系统相比,具有灵活性高、可控性强、寿命长等优点,因此得到了愈来愈广泛的应用。例如,数字录音电话、语音信箱、电子留言簿、发声字典、多媒体查询系统以及各种电子玩具等。在这类语音编码系统中,对编码实时性要求不高,往往不一定要求进行实时编码,但要求有较高的数据压缩率,以降低数字存储的容量。对译码器则要求算法尽量简单,成本尽量低,能够实时译码或基本实时译码。例如,在发声字典和多媒体查询系统中,编码器的需要量很

少,只要在厂家或信息中心装备即可,其复杂度可以较高,也不必完成实时编码。其译码器就不同了,用量很大,不但要求成本低,还要求能实时或基本实时译码,以减少响应时延。

数字电话通信系统与模拟电话通信系统相比,具有抗干扰能力强、保密性好、易于集成化等一系列优点,故此得到迅速发展和广泛应用。在数字电话通信系统中,一般都要求实时编码和译码,同时对算法复杂度、抗误码能力等都有一定要求。

1.2 语音编码的分类

1.2.1 传统的分类方法

语音编码按照传统的分类方法通常分为三类:波形编码、参数编码和混合编码。

波形编码是将时间域或频率域或变换域信号直接编码为数字信号,力求使重建语音波形保持原始语音信号的波形形状。波形编码具有语音质量好、抗噪声性能强等优点。其缺点是所需用的编码速率高,一般在 $64\sim 16\text{ kb/s}$ 之间。脉冲编码调制(PCM)、增量调制(ΔM 或 DM)、自适应增量调制(ADM)、自适应差分脉码调制(ADPCM)等,都属于波形编码。波形编码当其编码速率进一步降低时,其语音质量等性能指标下降很快。

参数编码又称为声源编码或声码器,有时又称为分析-综合编码,它是将信源信号在频域或其他变换域提取特征参数,然后对这些特征参数进行编码和传输;在译码端再将收到的数字信号译成特征参数,根据这些特征参数重建语音信号。参数编码是通过语音信号特征参数的提取和编码,力求使重建语音信号具有尽可能高的可懂度,即保持原语音信号的语意,但重建语音信号的波形与原语音信号波形却相差甚远。参数编码的优点是实现低速率语音编码,其编码速率可低至 2.4 kb/s 以下。其缺点是语音质量差,自然度较低,即使是熟人一般也听不出讲话人是谁。此外,参数编码的坚韧性也不够好。LPC-10 和 LPC-10e 声码器就属于参数编码。

计算机技术的发展为语音编码技术的发展提供了强有力的工具,大规模和超大规模集成电路的出现则为各种语音编码算法的实现提供了基础。最近二十几年来,语音编码技术取得了许多突破性进展,产生了一系列新一代的编码算法,这就是混合编码。混合编码将波形编码和参数编码结合起来,克服了波形编码和参数编码的缺点,吸收了它们的长处,在 $16\sim 4\text{ kb/s}$ 速率上能够得到高质量的合成语音。混合编码技术在现代通信系统中得到广泛应用。多脉冲激励线性预测(MPE-LP)编码、规则脉冲激励线性预测(RPE-LP)编码和码激励线性预测(CELP)编码等,

都属于混合编码。混合编码技术仍然处于迅速发展之中,目前仍有许多新的编码算法不断出现。

1.2.2 按照编码速率的分类方法

按照编码后传输所需的数据速率来分,可将语音编码划分为五类:

- (1) 高速率语音编码,其编码速率为 32 kb/s 以上。
- (2) 中高速率语音编码,其编码速率为 32~ 16 kb/s。
- (3) 中速率语音编码,其编码速率为 16~ 4.8 kb/s。
- (4) 低速率语音编码,其编码速率为 4.8~ 1.2 kb/s。
- (5) 极低速率语音编码,其编码速率为 1.2 kb/s 以下。

编码速率在 16 kb/s 以下的语音编码,通常称为语音压缩编码。低速率语音编码和极低速率语音编码是现代语音编码技术研究和发展的一个重要方向。

1.2.3 按照被编码信号所在的域的分类方法

按照被编码信号所在的域可以将语音编码划分为时域编码和频域编码。

时域编码是指对语音的时间域信号直接进行编码。PCM、ADM、ADPCM 等都属于时域编码。

频域编码是指对语音的频率域信号进行编码。子带编码(SBC)就是一种频域编码。

1.2.4 按照编码所使用的主要技术的分类方法

按照编码所使用的主要技术,可以将语音编码划分为线性预测编码、自适应编码、矢量编码和变换域编码等。

以线性预测作为主要技术基础的编码称为线性预测编码。混合激励线性预测声码器(MELPC)、多脉冲激励线性预测声码器(MPE-LPC)等,都属于线性预测编码。

采用自适应技术的编码称为自适应编码。自适应增量调制(ADM)、自适应差分脉码调制(ADPCM)等,都属于自适应编码。

采用矢量量化技术的编码称为矢量编码。由于矢量量化技术是一种高效数据压缩技术,所以在现代语音编码中得到广泛应用。

目前的许多编码算法,尤其是低速率和极低速率语音编码,大部分都采用矢量量化技术,都属于矢量编码。

变换域编码又称为变换编码,它首先采用某种正交变换将语音信号变换到变换域,然后再对变换后的信号进行编码。离散余弦变换(DCT)编码、离散傅里叶变换(DFT)编码等,都属于变换域编码。

应该指出,广义来说,子带编码等频域编码也是一种变换域编码,但其变换通常并非采用正交变换,因此,一般不把它们归为变换域编码。而变换域编码中的DFT变换编码也可称为频域编码,但其变换属于正交变换,故一般仍将其归为变换域编码。

1.2.5 按照编码算法是否依赖于某种模型假定的分类方法

按照编码算法是否依赖于某种模型的假定,可以将语音编码划分为基于模型的编码和不基于模型的编码两大类。有的文献资料又称之为模型编码和非模型编码。

不基于模型的编码或非模型编码,主要是指波形编码(包括变换编码)。这类编码是以波形逼近为原则,直接对时域波形或在变换域进行编码。其重建语音的质量好,但是编码所需速率较高。

基于模型的编码是指那些以某种模型的假定为基础的编码。这里的模型假定包括语音产生模型和听觉模型两方面。基于语音产生模型的编码有很多种,通道声码器、同态声码器、相位声码器、共振峰声码器、基于全极点声道模型的线性预测声码器(LPC)和许多由LPC改进而得到的混合编码方法,都是基于模型的编码。由于通道声码器、同态声码器、相位声码器和共振峰声码器等所用的编码方法已成为历史,现在很少应用。因此,我们将只重点介绍LPC声码器及由其改进而得到的各种混合编码算法。

1.2.6 按照被编码信号的属性的分类方法

按照被编码信号的属性可分为语音编码和声频编码。所谓声频编码,是指那些非语音的声音信号的编码,例如各种乐器发出的音乐声、鸟叫声、流水声等的编码。声频编码又称为音频编码,是语音编码的一个重要分支。

1.2.7 按照编码速率是否固定的分类方法

按照编码速率是否固定,可以将语音编码分为固定速率语音编码和变速率语音编码两类。

在语音编码过程中,始终保持输出数据速率固定的称为定速率语音编码。PCM、ADPCM、MPELP、CELP等大多数语音编码都是定速率语音编码。

在语音编码过程中,根据输入信号的不同情况而改变输出数据速率的称为变速率语音编码。第三代移动通信中使用的QCELP、EVRC等,就是变速率语音编码。

1.2.8 按照编码出现和应用时间以及技术发展情况的分类方法

按照编码出现和应用的时间以及技术发展情况,可以将语音编码分为经典语音编码和现代语音编码两类。20世纪70年代末以前出现和应用的语音编码,其技术发展已经相当成熟和完善,例如PCM,称为经典的语音编码。20世纪80年代以后出现和应用、目前在技术上仍在继续发展和完善的语音编码,例如CELP、QCELP等,称为现代语音编码。

此外,还有许多其他分类方法,这里不再赘述。

1.3 衡量语音编码性能的主要指标

语音编码所要解决的基本问题,是在给定的编码速率条件下,如何得到尽可能好的重建语音质量(或称编码质量),同时应尽可能减少编译码算法的复杂度和时延,并使编译码系统有较好的坚韧性;或者是在给定编码质量、编译码复杂度和时延以及坚韧性要求的条件下,如何尽可能降低语音编码所需的速率等。这五个方面的要求,就是衡量语音编码性能的主要指标。应该指出的是,这五个方面的要求(或称五个质量指标)有时是互相矛盾的,它们之间有着紧密的联系,是互相影响的,在不同的应用中,对各个指标要求的侧重点也有所不同。

1.3.1 编码质量

语音编码质量是衡量语音编码优劣的关键指标之一。评价语音编码质量的方法很多,归纳起来可以分成两类,即客观评定法和主观评定法。

1. 客观评定法

客观评定法用客观测量的手段来评定语音编码质量,常用的方法有信噪比、加权信噪比、平均分信噪比等。它们都是建立在度量均方误差的基础上,其特点是计算简单,但不能完全反映人对语音质量的主观感觉,这个问题对于速率在16 kb/s以下的中、低速率和极低速率语音编码尤为突出,因此客观评定法大多应用于较高速率的波形编码。

2. 主观评定法

主观评定法符合人类听话时对语音质量的感觉,目前得到了广泛应用。常用的方法有平均意见得分(meant opinion score, MOS)、诊断押韵测试(diagnostic rhyme test, DRT),判断满意度测量(diagnostic acceptability measure, DAM)等。

MOS得分采用五级评分标准,如表1.3.1所示。参加测试的实验者,在听完所

测语音后,从这五个等级中选择其中某一级作为他对所测语音质量的评定。全体实验者的平均分就是所测语音质量的 MOS 分。由于主观上和客观上的种种原因,每次测试所得的 MOS 分会有波动。为了减小波动的方差,除了参加测试的实验者人数要足够多之外,所测语音材料也要足够丰富,测试环境也应尽量保持相同。

表 1.3.1 MOS 判分五级标准及相应的描述该级语音质量的形容词

MOS 判分	质量级别	失真级别
5	优	不察觉
4	良	刚有察觉
3	可	有察觉且稍觉可厌
2	差	明显察觉且可厌但可忍受
1	坏	不可忍受

在数字语音通信中,通常认为 MOS 分 4.0~ 4.5 分为高质量数字化语音,达到长途电话网的质量要求,也常称之为网络质量。MOS 分 3.5 分左右称作通信质量,这时能感到重建话音质量有所下降,但不妨碍正常通话,可以满足多数话音通信系统使用要求。MOS 分 3.0 分以下常称合成语音质量,是指一些声码器合成的语音所能达到的质量。它一般具有足够高的可懂度,但自然度及讲话人的确认等方面的性能较差。

诊断押韵测试(DRT)是反映话音清晰度或可懂度的一种测试方法,它主要用于低速率语音编码的质量测试,因为这时可懂度已成为主要问题。这种测试方法使用若干对(通常为 96 对)同韵母的字进行测试,例如中文的“为”和“费”,英文的“fast”和“vast”等。测试中,让受试者每次听到一对韵字中的某个音,然后让他判断所听到的音是哪一个字,全体实验者判断正确的百分比就是 DRT 得分。通常认为 DRT 为 95% 以上时清晰度为优,85%~ 94% 为良,75%~ 84% 为中,65%~ 75% 为差,而 65% 以下为不可接受。在实际通话中,清晰度为 50% 时,整句的可懂度大约为 80%,这是因为整句中具有较高的多余度,即使个别字听不清楚,人们也能理解整句话的意思。当清晰度为 90% 时,整句话的可懂度已接近 100%,所以对于低速率语音编码,一般要求其清晰度能达到 90% 或以上。

判断满意度测量(DAM)是对话音质量的综合评估,它是在多种条件下对话音质量可接受程度的一种度量,也采用百分比评分。详细说明请参考有关文献资料,这里不再赘述。

1.3.2 编码速率

编码速率可以用“比特/秒”(b/s)来度量,它代表了编码的总速率,一般用 V 表示。编码速率也可以用“比特/样点”(b/p)表示,它代表了平均每个语音样点用

多少比特编码,一般用 R 表示。 V 和 R 可以通过取样频率联系起来:

$$V = Rf_s$$

其中的取样频率 f_s 通常根据 Nyquist 定理由信号带宽决定。目前国际上数字声频信号常用的带宽及取样频率由表 1.3.2 给出。

表 1.3.2 数字声频带带宽及取样频率

信号类型	取样频率/kHz	带宽/kHz	频率范围/Hz
电话	8	3.2	300~ 3400
电话会议	16	7	50~ 7000
激光唱片(CD)	44.1	20	20~ 20000
数字声广播及数字声磁带	48	20	20~ 20000

显然,平均每样点比特数 R 越高,语音波形或参数量化就越精细,语音质量也就越好,相应地对传输带宽或存储容量的要求也就越高。在波形编码中,为了获得高质量的重建语音,一般取 $R \geq 2(\Delta M$ 除外),而在参数编码中 R 可以低到 0.25 甚至 0.1 以下。当然在后一种情况下重建语音只能保持一定的可懂度,而反映讲话人特征的信息和语音的自然度就会有较大的损失。

1.3.3 编译码复杂程度

编码和译码算法的复杂程度同语音编码的语音质量有非常密切的关系。在同样数码率的情况下,采用复杂一些的算法将会获得更好的语音质量;而对于相同的话音质量,采用复杂一些的算法能够降低编码所需的速率。编解码算法的复杂程度同硬件的实现也有密切关系,它将决定硬件实现的复杂程度、体积、功耗以及成本等。目前在许多应用中都使用通用数字信号处理器芯片(DSP)实时实现各种语音编译码算法。它的好处是研制周期短,初次投资小。算法的复杂程度对 DSP 芯片的运算能力以及所需要的存储器容量都提出了一定的要求。运算能力可用每秒执行的百万条指令数(MIPS)衡量,存储器容量通常用千字或千字节来度量。算法越复杂则需要越高档的 DSP 芯片以及较大容量的存储器来实现,其成本、功耗目前都比较高。如果大量使用,则应该考虑开发和使用专用芯片。

1.3.4 编译码时延

增加算法的复杂程度可以提高语音编码质量,但往往也伴随着增加编译码的时延。在实时语音通信系统中,语音编译码的时延同线路传输时延的作用一样,对系统的通话质量有很大的影响。例如在卫星通信中,一跳传输时延约 0.5 s,讲话后再听到对方回答需要 1 s,已明显感到对方反应“迟钝”。如果时延再大,正常交谈都会发生困难。

时延影响通话质量的另一个原因是回声。在电话网中二/四线转换处,由于阻抗匹配不理想,甲端四线的收话音信号会泄漏到该端四线的发路径中去,它返回到乙端发话者,就形成回声。当时延比较小时,回声同话机侧音及房间交混回响声相混,因而感觉不到。当往返总时延超过约 100 ms 左右,发话者就能够从手机中听到自己的回声。如果回声传输路径损耗不够大,就会听到多次回声,从而严重影响通话质量。对于公用电话网,可能会有几次音频转接,也就是会有多次语音编译码,因此对单次语音编译码时延,通常要求不超过 5~ 10 ms。对于其他一些类型通信系统,例如卫星通信,由于传输时延或由于纠错编码引入的时延等已经比较大,再苛求语音编码的时延意义不大,它们主要对语音编码的话音质量、编码率以及抗误码性能等提出要求,在这些应用中,通常允许语音编码时延为几十毫秒到 100 ms。为了保证正常通话,当总时延超过 100 ms 时,一般都需要采取回声抵消或回声抑制等措施。

1.3.5 坚韧性

坚韧性(robustness)也是衡量语音编码性能优劣的重要指标之一。所谓坚韧性是指语音编码能够适应各种使用环境和条件,在较为不利的环境和条件下能正常工作。这包括以下几种情况:

- (1) 能够适应各种不同的讲话人的语音。
- (2) 能在较强的噪声背景下正常工作。
- (3) 在多级编码的情况下,语音编码质量不应有明显下降。
- (4) 容许一定误码,即在 10^{-2} ~ 10^{-3} 误码率时仍能提供易懂的重建语音。
- (5) 在部分数据丢失情况下,避免同步错乱等。

1.4 提高语音编码质量的基本途径

随着语音编码技术的发展,能够不断地在更低的速率上获得高质量编码效果。所使用的方法可以说是千变万化,但归纳起来,基本的依据或途径只有两条,这就是利用语音信号所存在的冗余度,以及利用人耳的听觉特性。

语音信号的冗余度主要源于两个方面,即语音信号幅度分布的非均匀性以及样点之间的相关性。语音信号的幅度统计特性同信号的带宽、录取时的声学条件以及进行统计的时间长度都有关系。对于高质量话筒、安静环境下录取的具有电话带宽(200~ 3400 Hz)的语音信号,其长时(几十秒以上)统计幅度特性接近于 Γ (gamma)分布,而短时(几至几十毫秒)统计所得幅度特性接近于高斯(Gaussian)分布。有时为了分析方便,长时统计特性也可以用双边指数或称为拉普拉斯(Laplacian)分布来近似。图 1.4.1 示出了这几种分布的概率密度函数(PDF)的图

形,而表 1.4.1 列出了它们的数学表达式。

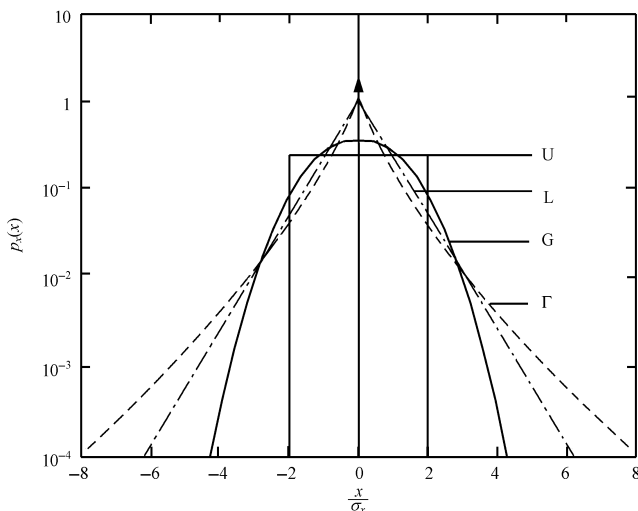


图 1.4.1 几种均值为 0 的概率密度函数的图形

表 1.4.1 均值为 0 的几种概率密度函数 $p(x)$

名称	符号	概率密度函数
均匀或矩形分布 (Uniform)	U	$\frac{1}{\Delta}, x \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ 0, 其他 $\sigma_x^2 = \frac{\Delta^2}{12}$
高斯或正态分布 (Gaussian)	G	$\frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{x^2}{2\sigma_x^2}\right\}$
拉普拉斯或双边指数分布 (Laplacian)	L	$\frac{1}{\sqrt{2\sigma_x^2}} \exp\left\{-\frac{\sqrt{2}}{\sigma_x^2} x \right\}$
Γ 分布 (Gamma)	Γ	$\frac{4\sqrt{3}}{\sqrt{8\pi\sigma_x^2} x } \exp\left\{-\frac{\sqrt{3}}{2\sigma_x^2} x \right\}$

无论根据长时或短时统计来看,语音信号都是小幅度出现的概率大,而大幅度出现的概率小。非均匀标量量化就是利用了语音信号的这一特点,使量化质量得到提高。CCITT G. 711 的 A 律或 μ 律 PCM 采用 8 b/p 的非均匀量化,达到了 12~13 b/p 均匀量化的效果。

语音样点间存在相关性是语音信号具有冗余度的另一个原因。语音是由肺呼

出的气流通过声门形成的激励信号激励声道再经唇口辐射而产生的。从信号处理的角度出发,可以把语音看成是由白色噪声或周期脉冲激励信号通过一个有色滤波波器所产生的。这一过程在时域上看,相当于使样点之间产生了相关性;从频域看,相当于给频谱加色,使原来的白色谱变成了非平坦的有色谱。此外,在浊音语音段,信号具有准周期性,其频谱含有谱线结构。因此,除了谱包络代表的短时相关性外,浊音段还具有长时相关性。利用语音信号的这些相关性可以有效地提高语音编码的质量。例如在时域中采用短时或长时线性预测,在变换域中对不同能量的谱分量分配不同的量化比特数目等,都能获得很好效果。

语音信号的另一个特点是短时平稳性。语音信号属于非平稳随机过程,但又具有短时平稳的特性。也就是说,语音的一些特征参数(例如短时能量、过零、相关系数等)在短小时内(几至几十毫秒)基本保持不变。相邻短时段参数变化也不会很大,即它们之间也存在着相关性,这是由于人的发音器官运动速度有限所决定的。利用语音信号的这一特点也可以大幅度压缩编码速率。例如在中、低速率编码中采用短时能量预测、LSP 参数的预测等可以进一步压缩编码速率。

关于人耳听觉特性在语音压缩编码中的应用主要有以下几方面

首先语音编码利用了人耳分辨率有限的特性。语音样点在幅度上是连续的,它的精确表示需要无穷多比特。实际上并不需要这样做,因为人耳对幅度的分辨能力是有限的。换句话说,语音信号对于人耳带有过多的信息,通过量化可以去除这些过多的信息。一般而言,线性量化中取 12~ 14 b/p 已听不出量化失真;若采用非线性量化,取 8 b/p 已能获得令人满意的效果;如果采用自适应量化,所需的比特还可以得到进一步压缩。

其次,语音编码,特别是中、低速率语音编码可以利用人耳的掩蔽效应来改善重建语音的质量。在编码过程中,通过采用非最小均方准则或其他方法,改变量化噪声的频谱形状,使得量化噪声在主观听觉上能够部分或全部被语音信号所屏蔽,从而达到提高语音编码主观质量的目的。在变换域变换中采用相应的方法,也能够达到同样的目的。

最后,利用人耳对某些失真不敏感的特性也能够达到压缩编码速率或提高主观语音质量的目的。线性预测声码器正是利用了人耳对相位失真不敏感的特点,索性不传送语音谱的相位信息,使编码速率能够压缩到 2.4 kb/s,甚至更低,而仍能保持很高的可懂度,而后滤波技术则是利用幅度谱的适度失真来降低量化噪声对语音质量的影响。

1.5 语音编码发展现状及今后的发展方向

1.5.1 语音编码的发展现状

语音编码的各种国际标准集中反映了语音编码技术发展的水平。标准的制定则是根据应用背景,对于编码质量、编码速率、编码时延、算法复杂程度以及坚韧性等几个因素进行综合权衡而做出的最优选择,以期在实际应用中获得最佳效益。表 1.5.1 列出了当前针对电话带宽语音编码而制定的一些主要国际标准以及它们的质量和范围。图 1.5.1 则更定量地描述了质量与编码速率之间的关系。虚线代表期望能达到的目标。

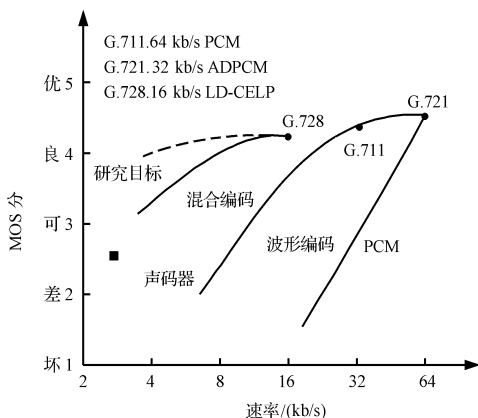


图 1.5.1 目前语音编码 MOS 分与编码速率之间的关系

从表 1.5.1 可以看到 20 世纪 80 年代以来语音编码技术迅速发展的情况,国际电报电话咨询委员会(CCITT)1972 年制定了关于 64 kb/s 脉码调制(PCM)语音编码标准的 G.711 建议。由于受到计算机技术、编码技术以及集成电路技术发展水平的限制,其后停顿了较长一段时间,直到 1984 年才公布了关于 32 kb/s 自适应差分脉码调制(ADPCM)编码标准的 G.721 建议,并且在 1986 年根据运行中出现的问题,对算法做了部分修改。这两个标准都达到了进入公用电话网的质量。码激励线性预测(CELP)编码方法提出后,16~ 4 kb/s 中、低速率语音编码的语音质量有明显提高。CCITT 从 1988 年开始研究符合进入公用网要求的 16 kb/s 语音编码标准,对该速率的算法提出如表 1.5.2 所示的一些主要要求。

表 1.5.1 目前国际上一些电话语音编码标准算法的性能及主要应用

编码标准 公布年代	CCITT G. 711 1972	CCITT G. 721 1984~ 1986	CCITT G. 728 1992	GSM 1988	CTIA 1989	NSA 1989	NSA 1982
比特率 算法名称	64 kb/s PCM	32 kb/s ADPCM	16 kb/s LD-CELP	13 kb/s RPE/LTP	8 kb/s VSELP	4.8 kb/s CELP	2.4 kb/s LPC
应用领域	公用电话网	公用网	公用网	数字移动通信	语音邮件	保密	话音
语音质量 五分制 MOS 分	4.3	4.1	4.1	3.7	3.8	3.2	2.5
DRT	95	94	94	—	93	93	90
DAM	73	68	70	—	68	64	53

表 1.5.2 中的 qdu 是一种量化噪声单位, G. 711PCM 一次编解码产生的量化噪声定义为单位 qdu。量化失真 ≤ 4 qdu 大体相当于 MOS 分为 4.0 以上。此表中同时要求时延 ≤ 5 ms 和量化失真 ≤ 4 qdu 是十分苛刻的, 当时没有一种现成的算法能够满足。有些算法量化失真接近要求, 但采用前向分帧处理, 时延远大于 5 ms。经过一年多的工作, CCITT 选定了 AT&T Bell 实验室提出的 16 kb/s 短时延线性预测编码方案(LD-CELP), 并经过进一步的研究和优化, 于 1992 年和 1993 年分别通过了浮点和定点算法(G. 728)。该算法采用了后向自适应算法, 即从量化后的语音信号中提取 LPC 参数及增益系数, 其时延小于 2 ms。经过测试, 该算法全面满足表 1.5.2 提出的各项要求。

表 1.5.2 CCITT 对 16 kb/s 高质量短时延语音编码算法的一些主要要求

参数	CCITT 要求	最后目标
编码延时	< 5 ms	< 2 ms
语音质量 无误码 误码为 10^{-3} 误码为 10^{-2}	量化失真 ≤ 4 qdu 优于 G. 721 32 kb/s ADPCM 优于 G. 721 32 kb/s ADPCM	—
多级编译码	3 次非同步转接 量化失真 ≤ 14 qdu	同步转接无量化失真累积
传送音频信令	能传送 DTMF 信号 (双音多频信令)	—

为了构成统一的移动通信系统, 从 1982 年开始, 西欧着手研究新一代数字移动通信体制, 并于 1988 年公布了标准, 称为 GSM 标准。其中语音编码采用了速率

13 kb/s 具有长时预测的规则脉冲激励的线性预测(RPE-LTP)方案。它的特点是算法简单、话音质量达到通信等级、抗误码性能好,加纠错保护后(速率达到 22.8 kb/s)能在 10^{-1} 突发性信道误码情况下工作,适用于数字移动通信。

与此同时北美、日本也在研究自己的数字移动通信体制。1989 年北美蜂窝电话工业组织(CTIA)公布了北美数字移动通信标准(IS-54 标准)。为了使原有模拟调频信道能扩容 3 倍,语音编码采用了 8 kb/s 矢量和激励线性预测(VSELP)方案。该算法属于 CELP 类型,只不过采用了自适应长时码书以及两个结构化的随机激励码书。比起经典的 CELP,VSELP 运算量减少很多,并能用定点算法实现,有利于硬件的简化。其质量也达到通信等级,并且比 GSM 的 13 kb/s RPE-LTP 编码器的质量略高。

1982 年美国安全局(NSA)公布了 2.4 kb/s 的 LPC-10 声码器标准(FS-1015 标准),采用 10 阶 LPC 预测器。它能给出 DRT 为 90% 的可懂度,因此在保密通信网中得到了广泛的应用。它的问题是合成语音自然度差且抗环境噪声能力差。为了提高保密通信网的话音质量,NSA 从 1985 年开始组织研究新一代声码器,并于 1989 年公布了 4.8 kb/s CELP 标准(FS-1016 标准)。在这个方案中采用了自适应长时码书和叠接稀疏随机码书,所需计算量也比经典 CELP 小。它具有比 LPC-10 好得多的自然度及抗环境噪声能力。1993 年美国国防部语音信号数字处理协会(DDVPC)开始选择新的国防部 2.4 kb/s 语音编码标准,并于 1996 年 3 月选用 TI 公司和 ASP 公司联合提出的混合激励线性预测(mixed excitation linear prediction, MELP)声码器取代旧标准 FS-1015,于 1998 年年底正式公布这个新的标准。这个 MELP 声码器的取样频率也是 8 kHz,每 180 个样点为一帧,帧长为 225 ms,每帧量化 54 bit,总的速率为 2.4 kb/s。MELP 声码器采用混合激励、非周期脉冲、自适应增强、脉冲整形滤波和傅氏级数幅度值等五大技术,使得合成语音能更好地接近自然语音。

除了上述提到的一些标准外,其他一些国家与组织也提出了一些标准算法。例如日本的数字移动通信标准(6.7 kb/s VSELP),国际海事卫星组织(INMARSAT)1990 年公布的 4.15 kb/s 改进型多带激励(IMBE)语音编码标准等。这些国家和组织还正在研究进一步降低速率的新的算法和标准,例如北美、西欧和日本正在制定将各自的数字移动通信语音编码速率降低一半的新的算法标准。CCITT 目前已经制定高质量短时延的 8 kb/s 语音编码标准,并开始制定 4 kb/s 乃至 2 kb/s 的高质量短时延的语音编码标准。

CELP 方案在 16~4 kb/s 速率上获得了巨大的成功。当速率低于 4 kb/s 时,由于码书容量变得太小,不能很好地代表预测余量信号,性能会很快下降。研究者普遍认为必须寻求新的途径来解决 4 kb/s 以下速率的高质量语音编码的问题,多带激励(MBE)编码和混合激励线性预测(MELP)编码已能实现 2.4 kb/s 高音质

语音编码。2 kb/s 以下速率的高音质语音编码是目前语音编码研究的主要课题之一。短时延编码也是目前语音编码的一个趋向,现在已经解决了 8 kb/s 低时延高质量的课题。更低速率、更高质量及短时延语音编码的算法研究已成为研究者当前的重要课题之一。

1.5.2 现代语音编码的发展趋势及方向

目前,由于通信的发展迫切要求各种技术的标准化,因此,语音编码技术的标准化已经取得了很大进展。ITU-T (CCITT) 制定了 64 kb/s、32 kb/s、16 kb/s、8 kb/s 等速率的语音编码标准。8~4 kb/s 语音编码也有了一些地区标准,这一速率段的语音编码国际标准以及 4 kb/s 以下速率的语音编码国际标准的制定,将是各国竞相研究开发的竞争热点。所以,低速率、极低速率语音编码将是今后语音编码研究和发展的一个重要方向之一。

此外,近年来在数字声广播、消费电子技术以及电话会议迅速发展的影响下,语音编码也在朝着宽频带高音质声频编码的方向发展。这是语音编码技术目前一个很重要的发展动向,受到人们的广泛关注。其应用领域主要是数字声广播、电视会议系统、家电及娱乐电子技术。由于受到社会需求的强大推动,这种编码也是一个相当重要的发展方向。

第三代移动通信(3G)和第四代移动通信(4G)的语音编码技术是语音编码另一个备受关注的研究课题。3G 的三大主流技术 TD-SCDMA、WCDMA 和 CDMA2000,都采用了变速率语音编码技术。TD-SCDMA 拟采用自适应多速率(AMR)语音编码;WCDMA 拟采用 AMR 语音编码,同时将可选模式声码器(SMV)作为备用技术;CDMA2000 选用变速率 CELP(QCELP)声码器或增强型变速率编码器(EVRC)。此外,在数字卫星通信等系统中,也有的采用了变速率语音编码技术。变速率语音编码由于能够在保证重建信号质量的条件下进一步压缩编码速率,其应用前景十分广阔,是语音及音频编码另一个重要的发展方向。

近年来随着 DSP 芯片技术、语音学、语义学、神经生理学和心理学的迅速发展和广泛应用,发展起来一种利用人类听觉系统特性的感知编码技术,在音频压缩编码中受到广泛关注,应用愈来愈广。这是一个重要的发展方向。

国际标准化组织(ISO)和国际电工技术委员会(IEC)组织运动图像专家组(MPEG)于 1992 年 11 月制定了关于视频和音频信号压缩的国际标准 ISO/IEC 11172,即 MPEG-1 标准。这个标准有系统、视频和音频三大主要部分,最大音频比特率为 1.856 Mb/s。MPEG-1 中有关音频压缩的标准,即 11172-3,已经成功应用在 VCD、CD-ROM、ISDN、数字音频广播以及视频游戏等领域中,它支持每声道为 224~32 kb/s 的 32 kHz、44.1 kHz 和 48 kHz 的 PCM 数据。现在 MPEG-2、MPEG-4 等的音频层编码标准也已经制定出来,目前非常流行的 MP3、MP4 等,