



普通高等教育“十二五”规划教材

21 世纪高等学校通信类规划教材

数字语音编码

姚天任 编著



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

普通高等教育“十二五”规划教材

21 世纪高等学校通信类规划教材

数字语音编码

姚天任 编著

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书系统介绍数字语音编码的基本理论、技术和分析方法, 主要包括: 语音编码发展概况, 语音信号和人的听觉特性, 数字语音编码基本技术, 语音波形编码, 声码器, 合成分析线性预测编码。全书突出介绍多脉冲激励、规则脉冲激励和码激励线性预测编码, 其中又以应用最广泛和最有发展前途的码激励线性预测编码为重点。书中选择 GSM 13 kb/s、G.722、FED-STD1016 和 G.728 等语音编码标准作为实例进行详细分析。书中安排了一定数量的习题。本书可作为高等学校信息与通信工程学科各专业高年级本科生或研究生的教材, 也可供相关领域的科学研究人员和工程技术人员参考。

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究。

图书在版编目(CIP)数据

数字语音编码 / 姚天任编著. —北京: 电子工业出版社, 2011.12

21 世纪高等学校通信类规划教材

ISBN 978-7-121-15339-6

I. ①数… II. ①姚… III. ①语音数据处理—编码—高等学校—教材 IV. ①TN912.3

中国版本图书馆 CIP 数据核字 (2011) 第 246968 号

责任编辑: 韩同平 特约编辑: 林宏峰

印 刷: 北京市李史山胶印厂

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 20.25 字数: 520 千字

印 次: 2011 年 12 月第 1 次印刷

印 数: 3 000 册 定价: 45.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

出版说明

教材建设是高等学校组织教学和进行学科建设的重要内容。

为贯彻落实教育部、财政部《关于实施高等学校本科教学质量与教学改革工程的意见》，根据教育部《关于进一步深化本科教学改革全面提高教学质量的若干意见》（教高〔2007〕1号）精神，为全面落实“质量工程”的具体工作，把课程、教材建设与资源共享提上新的高度，电子工业出版社以教育部普通高等教育“十二五”国家级规划教材的申报和出版为契机，充分发挥在电子信息领域的专业出版优势和独树一帜的品牌影响力，对2005年起已经出版的通信类教材，重新进行了整合，适时推出了这套“21世纪高等学校通信类规划教材”。

与以往出版的同类教材相比，这套教材具有以下特点：

（1）专业特色鲜明：以重点院校本科通信类专业的专业课程教材为主线，兼顾其他相关专业的通信类专业课程。

（2）系统性突出：本套规划教材覆盖了本科通信类专业的专业基础课、专业方向课及专业选修课，形成了一个完整的教材系列。同时注意教材之间课程内容的合理划分与衔接，层次分明，重点突出，各高校可以根据需要组合选用。我们的目的是为通信类课程打造一套全方位解决方案。

（3）体系、内容新颖：整个知识点建立在“高”、“新”平台上。基本理论阐述精练，深入浅出，便于自学；注意吸收新理论、新技术成果在人才培养中的作用；加强实践性与应用性，结合实例进行讲解。

（4）配套教学支持：多数教材配有教学课件（电子教案），部分重要课程配套出版教学辅导书或实验教材。

（5）质量保证：为从根本上保证教材的先进性和教学适用性，本套教材特别吸纳或整合了以下优秀资源：

①多数教材来源于优秀教材的修订和再版，包括教育部“十五”和“十一五”等国家级规划教材、国家级或省部级获奖教材、历届全国统编教材等。

②多数作者为著名教材作者、国家级/省级教学名师、国家级/省级优秀教学团队负责人。

③其他优秀教学资源，如国家级/省级精品课程、国家级实验教学示范中心、国家电工电子教学基地等组织编写的、体现相应课程或课程群特色的配套教材。

本套教材可作为高等学校通信专业及相关专业的本科生或研究生教材，也可供通信领域的有关专业人员学习参考。

为作好本套教材的出版工作，我们聘请了多位国内通信教育领域的著名教授作为教材顾问，并聘请了清华大学、东南大学、上海交通大学、北京交通大学、北京邮电大学、西安电子科技大学、电子科技大学等著名高校电子信息学院（系）的院长（系主任）成立教材编委会，从根本上保证了教材的高质量。在此对他们的辛勤工作表示衷心的感谢。

今后，我们将进一步加强同各高校教师的密切联系和合作，广泛听取一线教师对教材的反馈意见和建议，以便使我们的教材出版工作做得更好。

我们坚信：一流的教学队伍，一定有一流的教学理念和方法；一流的教学内容，需要配备一流的教材，从而体现一流的教学管理和教学质量。

《21世纪高等学校通信类规划教材》顾问委员

(按姓名音序排列)

迟惠生 (北京大学)	程时昕 (东南大学)
冯重熙 (清华大学)	李承恕 (北京交通大学)
吴伟陵 (北京邮电大学)	吴诗其 (电子科技大学)
谢希仁 (解放军理工大学)	袁保宗 (北京交通大学)

《21世纪高等学校通信类规划教材》编审委员

(按姓名音序排列)

主任委员 樊昌信 (西安电子科技大学)

副主任委员

顾畹仪 (北京邮电大学)	李建东 (西安电子科技大学)
彭启琮 (电子科技大学)	王金龙 (解放军理工大学)
王希勤 (清华大学)	文宏武 (电子工业出版社)
吴镇扬 (东南大学)	张思东 (北京交通大学)

委员

安建平 (北京理工大学)	鲍长春 (北京工业大学)
陈咏恩 (同济大学)	邓建国 (西安交通大学)
段哲民 (西北工业大学)	樊昌信 (西安电子科技大学)
范平志 (西南交通大学)	方 勇 (上海大学)
鄯广增 (南京邮电大学)	顾畹仪 (北京邮电大学)
顾学迈 (哈尔滨工业大学)	康 健 (吉林大学)
李建东 (西安电子科技大学)	李晓峰 (电子科技大学)
刘 琚 (山东大学)	彭启琮 (电子科技大学)
仇佩亮 (浙江大学)	唐朝京 (国防科技大学)
唐向宏 (杭州电子科技大学)	田宝玉 (北京邮电大学)
王金龙 (解放军理工大学)	王希勤 (清华大学)
王祖林 (北京航空航天大学)	文宏武 (电子工业出版社)
韦 岗 (华南理工大学)	吴镇扬 (东南大学)
徐昌庆 (上海交通大学)	张德民 (重庆邮电学院)
张思东 (北京交通大学)	郑建生 (武汉大学)
朱光喜 (华中科技大学)	朱秀昌 (南京邮电大学)

投稿联系方式: 韩同平编辑 (010)88254525

E-mail: hantp@phei.com.cn

前 言

语音编码的目的是在保证语音质量（不同应用领域对语音质量有不同要求）的前提下，用尽可能少的二进制数对数字语音信号进行有效表示。语音编码技术有广泛应用领域，它是现代通信和信息工程领域不可或缺的关键技术，它已成为信息科学领域的研究热点。各种语音编码标准（国家的或地区性的或国际性的）的研究和制订，一直受到普遍重视，许多语音编码标准正在不同应用领域发挥重要作用。

在高等理工科学学校中，通信与信息学科类专业一般开设了语音编码课程，但相应的教材不是很多，本书正是为这类课程编写的。本书重点介绍有关语音编码的基本理论、技术和方法。为了给学生打下坚实基础，书中选择了若干典型语音编码标准详加分析。本书内容包括：第 1 章，语音编码概论。介绍语音编码的目的、性能指标、基本方法、网络通信对它的特殊要求、语音编码标准，以及语音编码的研究动态和展望。第 2 章，语音信号。介绍语音的产生过程及其数字模型、语音信号的特性、人的听觉系统及其功能。第 3 章，数字语音编码的基本技术。介绍语音信号的数字化方法、短时域处理技术、短时傅里叶变换技术、线性预测技术和矢量量化技术。第 4 章，语音波形编码。介绍脉冲编码调制、差分脉冲编码调制、增量调制、自适应量化、自适应差分脉冲编码调制，以及变换编码的原理和方法。第 5 章，声码器。介绍声码器的一般工作原理、通道声码器、共振峰声码器、同态声码器和线性预测声码器，重点介绍激励源模型和基音估计方法，详细分析 LPC-10 和 LPC-10e 两个语音编码标准。第 6 章，合成分析线性预测编码。介绍合成分析线性预测编码一般原理、多脉冲激励线性预测编码、规则脉冲激励线性预测编码、码激励线性预测编码，重点介绍码激励线性预测编码，包括搜索最优码矢的快速方法、预测器参数的计算方法、基音预测器和共振峰预测器的开环和闭环联合优化方法和基于维纳滤波的联合优化方法。对 GSM 13 kb/s、G.722、FED-STD1016 和 G.728 这 4 个语音编码标准进行了较详细分析。

为加深对语音编码基本概念的理解和对基本方法的掌握，每章末安排了一定数量的习题。

本教材参考学时为 48 学时，除打*号的内容外其余都是必学的内容，其中第 1 章和语音编码标准可以安排自学并组织课堂讨论。

本书可作为高年级本科生或研究生的教材，也可作为信息和通信学科领域中从事科学研究和工程应用的人员的参考书。

鉴于作者水平所限，书中难免存在缺点、不妥和错误之处，希望读者不吝赐教。

E-mail: ctianrenyao@mail.hust.edu.cn

作 者
于华中科技大学

目 录

第1章 概论	1
1.1 语音编码的目的	1
1.2 语音编码的性能指标	2
1.2.1 语音质量	2
1.2.2 比特率	3
1.2.3 延时	4
1.2.4 算法复杂度	5
1.3 语音编码的基本方法	6
1.4 网络通信应用中的语音编码问题	8
1.4.1 VoIP	8
1.4.2 嵌套编码和多模编码	8
1.4.3 信源和信道联合编码	9
1.5 语音编码标准	10
1.5.1 语音编码标准化组织	10
1.5.2 广泛应用的语音编码标准	11
1.5.3 语音编码标准的主要性能指标和应用领域	13
1.6 语音编码研究动态	21
1.7 展望	22
习题	22
第2章 语音信号	24
2.1 语音的产生过程及其模型	24
2.1.1 人类的语音器官	24
2.1.2 人类语音的产生过程	25
2.1.3 共振峰频率和基音频率	25
2.1.4 描述语音产生过程的离散时间信号模型	26
2.2 语音信号的特性	27
2.2.1 语音的声学特性	27
2.2.2 语音信号的波形	28
2.2.3 语音信号的统计特性	29
2.3 人类的听觉功能	32
2.3.1 人的听觉器官	32
2.3.2 耳蜗的构造	33
2.3.3 耳蜗的功能	34
2.3.4 等响度曲线	35
2.3.5 音调	37
2.3.6 临界频带	37
2.3.7 听觉掩蔽效应	40
习题	43

第3章 数字语音编码的基本技术	45
3.1 语音信号的数字化	45
3.1.1 反混叠失真滤波和取样	45
3.1.2 量化	46
3.1.3 量化定理	49
3.1.4 量化误差的统计特性	53
3.1.5 信号与量化误差的相关性	54
3.1.6 高频颤动	56
3.1.7 量化噪声谱整形技术	59
3.2 数字语音信号的短时域处理	62
3.2.1 短时能量	62
3.2.2 短时平均幅度	65
3.2.3 短时平均过零率	65
3.2.4 短时自相关函数	65
3.2.5 短时平均幅度差函数	66
3.2.6 语音段起止点判别	66
3.3 语音信号的短时傅里叶变换	67
3.3.1 窗函数对短时谱的影响	67
3.3.2 短时傅里叶变换的线性滤波实现	69
3.4 语音信号的线性预测分析	69
3.4.1 线性预测原理	69
3.4.2 线性预测的规范方程	71
3.4.3 求解线性预测方程的两种常用方法	72
3.4.4 Levinson-Durbin 算法	74
3.4.5 线性预测系数与反射系数的相互转换	77
3.4.6 协方差法	78
3.4.7 格形滤波器	81
3.4.8 Burg 算法	83
3.4.9 线性预测的 Schur 迭代算法	84
3.4.10 线性预测的其他参数表示	92
3.5 语音信号的矢量量化	96
3.5.1 矢量量化的基本原理	96
3.5.2 失真度	96
3.5.3 码本的设计	97
3.5.4 搜索码本的方法	101
习题	103
第4章 波形编码	107
4.1 概述	107
4.2 脉冲编码调制	108
4.2.1 均匀量化 PCM	108
4.2.2 误码引起的信号失真	109
4.2.3 非均匀量化 PCM	110

4.2.4	μ 律压扩器	111
4.2.5	A 律压扩器	113
4.2.6	μ 律 PCM 和 A 律 PCM 的编码规则	114
4.2.7	对数 PCM 与线性 PCM 间的转换	116
*4.2.8	实现 PCM 编解码的集成电路芯片	118
4.3	差分脉冲编码调制	119
4.3.1	DPCM 的基本原理	120
4.3.2	预测信号 $\hat{x}(n)$ 的产生方法	121
4.3.3	DPCM 的信号-量化噪声比	121
4.4	增量调制	123
4.4.1	增量调制工作原理	123
4.4.2	增量调制性能分析	124
4.5	自适应量化	125
4.5.1	自适应量化的基本原理	125
4.5.2	语音信号方差的估算	126
4.5.3	自适应量化对信噪比的改善	127
4.6	自适应差分脉冲编码调制	128
4.6.1	前馈自适应和反馈自适应预测	128
4.6.2	自适应差分脉冲编码调制的基本原理	129
4.6.3	ADPCM 系统的性能	130
4.7	子带编码	133
4.7.1	离散时间信号的减取样和增取样	133
4.7.2	短时傅里叶变换的滤波器组分析方法	137
4.7.3	子带编码原理方框图	139
4.7.4	双通道正交镜像滤波器组	139
4.7.5	宽带语音编码标准 G.722	142
4.8	变换编码	153
4.8.1	变换编码与短时傅里叶变换的关系	153
4.8.2	自适应变换编码原理	154
*4.8.3	酉变换的选择	156
4.8.4	变换系数的量化	156
4.8.5	比特分配	157
4.8.6	谱的参数化和变换编码的自适应	158
*4.8.7	自适应变换编码设计的典型参数举例	161
	习题	163
第 5 章	声码器	165
5.1	声码器工作原理	165
5.2	通道声码器和共振峰声码器	166
*5.3	同态声码器	169
5.3.1	卷积同态系统	169
5.3.2	复倒谱与倒谱	170
5.3.3	语音信号的倒谱分析	170

5.3.4	同态声码器原理	173
5.4	线性预测声码器	174
5.4.1	线性预测声码器的基本原理	174
5.4.2	线性预测参数的量化	175
5.4.3	线性预测声码器的实例	176
5.5	激励源模型	178
5.5.1	两状态激励模型	178
5.5.2	混合激励模型	179
5.5.3	脉冲和噪声混合比例的控制方法	180
5.5.4	脉冲位置抖动器	180
5.5.5	浊音激励的周期性和脉冲性	182
5.5.6	混合激励模型的进一步改进	183
*5.5.7	采用改进混合激励模型的声码器实例	188
5.5.8	残差激励线性预测	189
5.6	基音估计	191
5.6.1	基音检测的困难	192
5.6.2	基音检测算法的类型	192
5.6.3	基音检测算法	193
5.7	LPC-10 和 LPC-10e 声码器	196
5.7.1	LPC-10 声码器的编码原理	197
5.7.2	LPC-10 声码器的解码原理	200
*5.7.3	LPC-10e 声码器	202
	习题	204
第 6 章	合成分析线性预测编码	205
6.1	合成分析线性预测编码原理	205
6.2	多脉冲激励线性预测编码器	207
6.2.1	多脉冲激励的物理基础	207
6.2.2	多脉冲激励信号迭代优化计算过程	208
*6.2.3	多脉冲激励线性预测编码实例	211
6.2.4	9.6 kb/s MPLP 编码算法	212
6.3	规则脉冲激励线性预测编码	213
6.3.1	规则脉冲激励信号的优化过程	214
6.3.2	规则脉冲激励信号的优化	214
*6.3.3	激励信号优化算法的改进	216
6.4	GSM 13 kb/s 语音编码器	219
6.4.1	背景介绍	219
6.4.2	GSM 语音编码器工作原理	220
6.4.3	GSM 13 kb/s 语音编码器的解码原理	225
6.4.4	GSM 13 kb/s 语音编码器的主要性能	226
6.5	码激励线性预测编码技术	227
6.5.1	概述	227
6.5.2	CELP 编码器基本原理	228

6.5.3	传统 CELP 编码器的码本搜索算法	230
6.6	CELP 编码器搜索最优激励码矢的快速方法	234
6.6.1	矩阵奇异值分解搜索方法	234
6.6.2	频域搜索方法	235
6.6.3	自相关搜索方法	236
6.6.4	稀疏矢量搜索方法	237
6.6.5	谱分类搜索方法	240
6.6.6	利用代数码、短激励序列和树码加快码本搜索	241
6.7	CELP 编码器中预测器参数的计算	242
6.7.1	计算预测器参数的通用模型	242
6.7.2	共振峰预测器参数的计算	245
6.7.3	基音预测器参数的开环计算方法	246
6.7.4	修正协方差法	248
6.7.5	共振峰预测器和基音预测器的计算顺序	250
*6.7.6	长时合成滤波器的稳定性分析	252
6.7.7	基音周期(或基音滞后时间)的估计	253
6.7.8	基音预测器参数的闭环计算方法	255
*6.8	基音预测器与共振峰预测器的开环联合优化	257
6.8.1	组合结构联合优化方法	258
6.8.2	级联结构联合优化方法	258
6.8.3	开环联合优化滤波器的性能	261
6.8.4	开环联合优化中的稳定性问题	263
6.8.5	最小相位开环计算联合优化方法	263
6.9	基音预测器与共振峰预测器的闭环联合优化	266
*6.10	基于维纳滤波的联合优化方法	269
6.10.1	开环维纳滤波	269
6.10.2	闭环合成分析	271
6.11	FED-STD1016 语音编码标准	274
6.11.1	基本工作原理	274
6.11.2	码本构成方法	275
6.11.3	自适应码本编码表	280
6.11.4	固定码本和自适应码本的增益的编码	281
6.11.5	解码器中的后置滤波器	282
6.11.6	FED-STD1016 编码标准的主要性能指标	282
6.12	G.728 语音编码标准	283
6.12.1	G.728 的编码和解码原理	283
6.12.2	G.728 的解码器	292
6.12.3	G.728 编解码器中各种计算的时间安排	297
	习题	299
附录 A	英文缩略词中英文对照表	302
	参考文献	305

第1章 概 论

本章概述语音信号编码的基本问题，包括语音编码的目的、性能指标、基本方法、编码标准、应用领域，以及语音编码的研究方向。

1.1 语音编码的目的

语音编码的目的，是寻求语音信号的某种紧缩表示方法，以便能够在频带有限的有线信道和无线信道上有效传输，或在介质中存储这些信号时节约存储空间。说得更具体一些，语音编码的目的，就是用尽可能少的比特数以数字形式表示语音信号，同时保持要求的语音可懂度和语音质量。因此，语音编码又称为语音压缩编码或简称为语音压缩。

语音编码是公共交换电话网(the public switched telephone network, PSTN)、电视会议系统、数字蜂窝通信系统，以及 VoIP(the voice over internet protocol, 基于 IP 协议传输声音)等应用的基础。

推动语音编码技术发展的主要因素，是数字通信的发展，以及对降低比特率和节约带宽的普遍需要。语音编码总是在降低比特率和保持语音质量之间进行权衡，在某些应用中还必须考虑编解码的复杂度、编码延时，以及与误码和丢包有关的性能等其他因素。

语音编码的对象是数字化的语音信号和音频信号，语音信号是指人的语音经过数字化后得到的数字信号，音频信号是指人耳可闻的任何声波振动(频率范围约为 20~20 000 Hz)经过数字化后得到的信号。显然，音频信号是一个比语音信号含义更广的术语，它不仅包括语音信号，而且还包括音乐，以及人耳可闻的其他任何声音信号。

语音信号数字编码简称为语音编码。按照被编码信号(编码器的输入信号)所占据的频带宽度，语音编码可以分成三类：①窄带(或电话带宽)语音编码，输入信号频率范围为 200~3400 Hz；②宽带语音编码，输入信号频率范围为 50~7000Hz；③音频编码，输入信号频率范围为 20~20 000 Hz。通常情况下，人们习惯于不加区分地使用“语音编码”和“音频编码”两个术语；为了方便，在不需要加以区分的情况下，本书一律把语音和音频信号编码统称为语音编码。本书讨论的内容限于窄带和宽带语音信号的数字编码。

在不同应用领域，对语音和音频信号的频带宽度提出了不同的要求，如表 1.1 所示。表中同时列出了信号的频率范围和取样率。

通常提到语音编码，实际上总是包括编码和解码两个方面。而编码这个术语，可以是编码技术、编码设备或部件(编码器)、编码系统、编码算法等的简称，一般情况下没有必要对它们严格加以区分。

语音编码器是通信系统和多媒体系统的重要组成部分。图 1.1 所示的是数字通信系统的一般性方框图。图中的信源通常包含着冗余信息(信号取样值之间存在着相关性)，从提高带宽利用率

表 1.1 不同应用领域对语音和音频信号的频带要求

应用领域	取样率 (kHz)	频率范围 (Hz)	频带宽度 (kHz)
固定电话	8	200~3400	3.2
电话会议	16	50~7000	7
调幅无线广播	16	20~10 000	10
调频无线广播	32	20~15 000	15
光盘	44	20~20 000	20
数字磁带	48	20~20 000	20

的观点来看, 应该在传输前将冗余信息去掉; 由于这些冗余信息通常是不知道的, 因此不能用于克服信道噪声, 所以去掉它们有利无害。但是, 由于信源编码后的信号缺乏冗余信息, 所以被传送的信号非常容易受到信道噪声的影响。为了克服信道噪声的影响, 在信源编码后紧接着设置了信道编码, 信道编码有意识地加入数量可控的已知冗余信息, 而在接收端将这些已知冗余信息用于解码, 以达到减小信道噪声影响的目的。这样, 通过信源编码与信道编码的配合, 既去掉了未知的冗余信息, 又赋予了编码后信号抗噪声的能力。本书只讨论信源编码, 不涉及信道编码, 因此, 本书凡是未做特殊说明的所谓“编码”都是指信源编码。

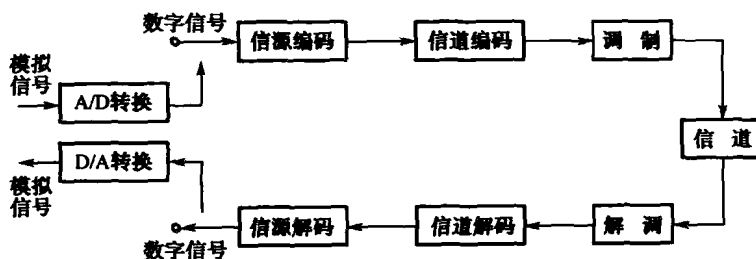


图 1.1 数字通信系统的方框图

在实际数字通信系统中, 除非专门产生数字信号作为信源, 发射端的信源一般都是模拟信号, 因而在信源编码之前需要采用模数转换器(A/D)将其转换成数字信号, 到了接收端再利用数模转换器(D/A)恢复成模拟信号, 提供给终端用户。

语音信号数字编码有非常广泛的应用领域, 其中包括: 通信, IP 电话, 视频会议, 消费电子, 电子存档, 数字同步的声音和数据(The Digital Simultaneous Voice and Data, DSVD), 以及大量的基于 PC 的游戏和多媒体应用。

1.2 语音编码的性能指标

语音编码的性能常用下列指标来衡量: 语音质量, 比特率, 延时, 算法复杂度。

1.2.1 语音质量

这里所说的语音质量是指解码器输出语音的质量。语音质量包括语音的可懂度、自然度, 以及说话人辨识等几个方面。评估编码语音质量的理想条件是: 输入语音清晰(无背景噪声), 无传输误码, 一次编码。重要的是, 在理想条件下, 解码后的语音听起来感觉到的质量如何。但是实际上很难满足这些理想条件, 因为: ①大多数实际应用环境都存在着大量背景噪声; ②传输过程中有误码(个别比特误码, 甚至整帧丢失); ③在会议电视等许多应用中, 常常需要桥接几个呼叫者, 以使每个呼叫者都能够听到其他呼叫者的谈话, 因此, 要求语音编解码器把每个呼叫者的码流解码后相加, 并将相加结果重新编码, 这就出现多次编码和解码的情况, 甚至出现不同编码标准进行级联的情况。在评估语音质量时, 所有这些情况都应该加以考虑。

评估编码语音质量有主观和客观两类度量标准。大多数客观度量标准是以分贝(dB)数表示的信噪比(SNR)为基础的, SNR 定义为

$$\text{SNR} = 10 \lg \frac{E[x^2(n)]}{E[e^2(n)]} = 10 \lg \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} [x(n) - \hat{x}(n)]^2} \quad (1.1)$$

式中, $x(n)$ 和 $\hat{x}(n)$ 分别是编码器的输入信号和解码器的输出信号, $e(n) = x(n) - \hat{x}(n)$ 是编码误差, $E[x^2(n)]$ 和 $E[e^2(n)]$ 分别是信号功率和噪声功率, N 是计算信号功率的语音段长度。由于语音信号是非平稳信号, 按式(1.1)计算得到的 SNR 是随语音段的不同而不同的, 因此常采用下式定义的平均分段信噪比

$$\text{SNR}_{\text{seg}} = \frac{1}{K} \sum_{k=1}^K \text{SNR}_k \quad (1.2)$$

式中, SNR_k 是用式(1.1)计算得到的第 k 帧语音信号的信噪比, K 是帧数。

基于 SNR 的度量主要用来说明重建语音与原始语音的波形的近似程度, 因此适合于波形编码。这种度量的计算简单, 而且可以根据它来确定重建语音的增益和延时。但是, 它没有考虑人耳的听觉感知特性, 因此, 在传输速率有限的应用中, 常常有可能出现信噪比高的重建语音不一定对应于听觉感知好的语音的情况, 即客观度量的语音质量与主观听觉感知的语音质量不一致的情况。所以在低比特率编码的应用中, 常常不使用客观度量方法而主要使用主观度量方法。

平均评价分(the mean opinion score, MOS)是应用最广的编码语音质量主观度量参数, 按照国际电信联盟(The International Telecommunication Union, ITU)和(北美)电信工业协会(The Telecommunication Industry Association, TIA)的规定, 它应该由 32~64 个(实际中往往只由 12~24 个)经过专业训练和“校对”的测试人员, 对某个编码语音的数百个语音记录的语音质量进行试听, 并用 5 级记分法进行评定, 然后将他们的评分进行平均, 便得到最后的 MOS 参数。MOS 参数主要反映的是人耳对语音的实际听觉感知效果, 从 1 分到 5 分所对应的语音质量等级分别是很差、差、一般、良和优。MOS ≥ 4 对应于网络或长途电话语音质量, MOS = 3.5~4 对应于蜂窝通信语音质量, MOS = 2.5~3.5 对应于合成语音质量。值得注意的是, 由不同人在不同时候测试得到的 MOS 参数是有差别的, 因此不应该把它作为对不同编码语音质量进行比较的绝对标准。

除了 MOS 参数外, 编码语音的主观质量参数还有以下几个:

(1) 韵律识别测试(The Diagnostic Rhyme Test, DRT): 用一组发音相近的孤立字测试区别它们的能力。

(2) 满意程度(The Diagnostic Acceptability Measure, DAM): 类似于 MOS 参数, 但是采用了更复杂的加权平均算法。

(3) A/B 测试(A/B Test)。将目标算法与性能已知的算法进行比较。

1.2.2 比特率

比特率又称为码率, 是指对长为 1 s 的语音信号进行数字编码所用的比特数, 度量单位是每秒比特(b/s)。比特率反映语音被压缩的程度, 比特率越低说明语音压缩越多。

编码语音具有高的比特率, 意味着它要求占用宽的带宽。编码语音常常与其他数据公用

一个信道, 因此, 总是希望在保证满足解码语音质量要求的前提下, 使编码语音的最大比特率尽可能低, 以使占用的信道带宽尽量小。大多数语音编码器工作在固定的比特率, 而不管输入语音信号的特性。在多媒体应用的情况下, 由于编码语音与其他形式的信号共享同一信道, 因此最好让编码器的比特率是可变的。在语音和数据同时在同一个信道上传送的应用中, 可以在有语音信号的时段(语音活动期)使用一个固定比特率, 而在没有语音信号的时段(只有背景噪声的静默期)使用一个很低的比特率, 这就是所谓静默期压缩方法。采用静默期压缩方法时, 把静默期的比特率降到零(不传送任何比特)的做法是不可取的, 因为人们习惯于在静默期听到有低电平的背景噪声而不习惯于静默期死一般的寂静。如果在静默期没有传送任何比特, 那么, 在解码时应该人为地生成一种让人听了感到“舒服”的背景噪声加入到静默期中, 这就是所谓的“舒适噪声生成”方法。为了使语音活动段与静默段之间平滑过渡, 要求编码端和解码端准确同步; 由于在静默期没有任何比特传送, 因此这种同步有时是困难的。

静默期压缩方法首先需要准确检测语音的活动期, 这是由话音活动检测器(The Voice Activity Detector, VAD)来完成的。但是在背景噪声电平很大的情况下, 要准确区分语音与背景噪声是很困难的。如果 VAD 不能识别语音的开始, 那么开始的谈话就可能会被切断, 这就是所谓的语音“前端切断”(the front-end clipping)。前端切断将使语音的可懂度下降。

语音编码广泛应用于通信工程领域, 在该领域中常把语音质量分成如下 4 个等级:

- (1) 广播语音: 宽带(见表 1.1)高质量语音, 噪声和失真非常小以至于人耳感觉不出来。
- (2) 长途电话(即网络通信)语音: 窄带(见表 1.1)高质量语音, 信噪比不低于 30dB, 谐波失真不大于 2%~3%。

(3) 通信语音: 有较好的可懂度, 失真比长途电话语音的大, 质量也较差。

(4) 合成语音: 有一定可懂度, 但音质较差, 而且不能根据语音辨识说话人。

比特率与语音质量有密切关系, 一般的趋势是比特率越高解码语音质量越好。具体来说, 广播质量语音的比特率至少应该大于 64 kb/s; 长途电话质量语音的比特率为 8~64 kb/s, 取决于语音编码器的复杂度; 通信质量语音的比特率为 4.8~12 kb/s; 合成语音质量的比特率一般低于 4.8 kb/s。

图 1.2 形象地示意了语音质量与比特率的关系。

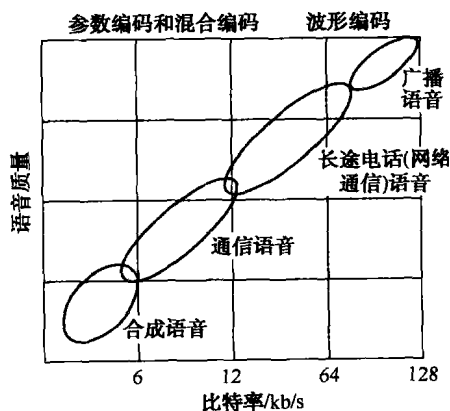


图 1.2 语音质量与比特率的关系

1.2.3 延时

语音编码系统的延时由运算延时、处理延时和通信延时等 3 部分组成。

许多低比特率语音编码器对数据的处理和传送是逐帧进行的, 此外, 有时为了能够对信号进行更充分的分析和处理, 需要使用一帧数据之前的若干数据(称为预做准备数据), 因此, 需要首先把一帧数据和其前面的若干数据加以缓存, 然后才能开始对这帧语音进行编码, 由此产生的延时称为运算延时。无论怎样改变编码和解码的方案, 运算延时总是无法减小的, 但是其他两种延时却能够通过硬件的改进来减小。

编码器的编码时间和解码器重建语音的时间的总和称为处理延时。处理延时取决于编码和解码的算法，以及实现编码器和解码器的硬件的速度。运算延时与处理延时之和称为单向编解码延时。

一帧数据从进入编码器输入端算起，直到由解码器输出端输出，中间不经过信道，总共需要的时间，称为通信延时。

上述的运算延时、处理延时和通信延时的总和称为编码系统的单向系统延时。在通信系统中，如果存在着由于阻抗不匹配而产生的回声，那么，编码延时将增大回声对语音信号的影响，并使语音质量下降。所以，在没有回声的情况下，最大单向系统延时甚至允许大到 400 ms。但是为了使通信流畅，最好能使最大单向系统延时不大于 200 ms。在有回声的情况下，要求最大单向系统延时不大于 25 ms。因此，通信系统中常常需要使用回声消除器。

在桥接等应用中，单向系统延时将加倍。而且多次编码也必将使语音质量下降。

ITU-T 制定的 G.114 建议标准，对于回声受到适当控制的单向系统延时提出的指标如下：

- (1) 0~150 ms: 大多数应用都满意的单向系统延时；
- (2) 150~400 ms: 是否满意取决于传输时间的影响；
- (3) 400 ms 以上: 一般的网络通信都不能满意。

图 1.3 是 ITU-T G.114 建议的标准中引用的一个实例，说明编码的单向系统延时对语音质量的影响。从该图可以看出，0~150 ms 的延时对应于曲线的平坦部分，有时把对语音质量很满意的延时范围扩展到 200 ms；目前的考虑是把对语音质量很满意的延时范围进一步扩展到 250 ms，这相当于现有标准中“满意”这一档的中间的延时。

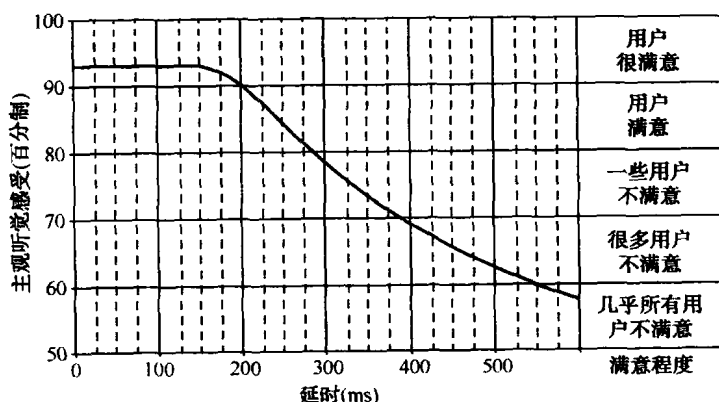


图 1.3 编码的单向系统延时对语音质量的影响

1.2.4 算法复杂度

语音编码的算法复杂度直接决定了硬件实现的复杂性、体积、功率消耗和价格等指标。大多数实时语音编码算法是在数字信号处理器(DSP)上实现的，因此，一个比较复杂的算法意味着要求功能更强大的 DSP 芯片。DSP 芯片的主要性能指标包括运算速度、随机存储器(RAM)和只读存储器(ROM)的容量等。在 DSP 芯片上集成较多的 RAM 和 ROM 意味着增加芯片的成本。DSP 芯片的运算速度用单位时间(每秒)执行百万条指令数(The Million Instructions Per Second, MIPS)来表示。低复杂度语音编码器需要的运算速度低于 15 MIPS，高复杂度语音编码器需要的运算速度高于 30 MIPS，中复杂度语音编码器需要的运算速度介于

这两个数值之间。

算法复杂度是设计或选择语音编码算法时需要重点考虑的因素。增加编码算法的复杂度意味着增加硬件实现的成本、增加硬件的功率消耗(对于便携式设备来说,意味着电池充电时间间隔缩短)。另一个需要重点考虑的因素是语音编码器的功率消耗和价格在整个系统中所占的比重。为了达到给定的比特率,如果能够把这个比重控制在 10%左右,可以尽量选择最好的编码器算法,因为各种不同编码器的功率消耗和价格之间的差别一般不足以改变这个比重。但是如果这个比重偏大,选择最合适的编码器就非常必要了。

1.3 语音编码的基本方法

语音压缩的物理基础:一是利用语音信号的冗余度,二是利用人耳的听觉特性。

减小语音信号冗余度主要从 4 个方面着手:①语音信号相邻取样值之间存在很强的相关性,可以采用适当的方法直接去掉;②语音信号中的浊音具有准周期性特点,这就允许利用语音合成模型来获得非常低的比特率;③在产生语音的过程中,人的发音器官(主要是声道)的形状随着时间的变化比较缓慢,因此可以将语音信号分帧进行短时分析和编码;④传输码值的分布不是均匀分布,因此可以采用各种不同的概率编码技术。

可用于语音编码的人耳听觉特性有:①人耳的听觉感知最主要的频率范围是 200~3400 Hz,频率分辨率约 2 Hz,因此在对语音信号进行量化时,可以根据这些数据去掉进入人耳的冗余信息;②人耳具有听觉掩蔽效应,例如强的声音可能会掩蔽弱的声音,其他因素也有可能引起听觉掩蔽,例如,频率相近的声音,发声时间相近的声音,噪声与纯音,等等;③人耳对不同频率的声音有不同灵敏度,一般对低频比对高频更灵敏,而语音的基音频率和共振峰频率主要处在低频段,这样,就可以对高频成分进行粗略编码;④人耳对于信号的相位不灵敏,因而可以去掉相位信息。

语音编码的基本方法分为以下三大类。

1. 波形编码

波形编码的基本思想是把语音信号当成普通的信号波形,不考虑语音本身的特点和人耳对语音听觉感知的特点,直接针对语音信号波形的每个取样值逐个地或若干个取样值成组地进行量化并编码,前者是标量量化,后者是矢量量化;解码时,力图尽可能准确地恢复原始语音信号的波形。如果用 $s(n)$ 和 $\hat{s}(n)$ 分别表示原始语音信号和解码器恢复(重建)的语音信号,用 $e(n) = s(n) - \hat{s}(n)$ 表示重建误差,那么,波形编码的准则是在给定的比特率下使 $e(n)$ 的功率最小。因此,信噪比(SNR)是波形编码的有用的性能指标。

可以看出,上述波形编码的基本方法实际上就是脉冲编码调制(The Pulse Code Modulation, PCM)。为了进一步去掉相邻取样值间的相关性,可以将相邻取样值相减得到差分信号,然后对差分信号进行编码,这就是差分脉冲编码调制(The Differential Pulse Code Modulation, DPCM)。其他重要的波形编码方法有自适应差分调制(The Adaptive Differential Modulation, ADM),自适应差分脉冲编码调制(The Adaptive Differential Pulse Code Modulation, ADPCM),自适应预测编码(The Adaptive Predictive Coding, APC),自适应子带编码(The Adaptive Sub-Band Coding, ASBC),以及自适应变换编码(The Adaptive Transform Coding, ATC)等。