

MODERN
COMMUNICATIONS

现代通信理论与技术丛书

Principles of Digital Speech Coding

数字语音编码原理

● 鲍长春 编著



西安电子科技大学出版社
<http://www.xduph.com>

内 容 简 介

本书以语音信号产生的源-系统模型为“绳”，系统而深入地介绍了低比特率数字语音编码的基本理论和基本原理。全书共 12 章。其中，第 1 章为语音编码概论；第 2~7 章介绍了低比特率数字语音编码的基本理论，包括语音信号的数字模型、语音信号的数字分析、语音信号的基音检测、语音信号的线性预测分析、语音信号的矢量量化方法和线性预测系数的有效量化方法；第 8~12 章介绍了国际上热衷研究并得到应用的五种低比特率数字语音编码原理和技术，包括码激励线性预测语音编码、波形内插语音编码、多带激励语音编码、正弦语音编码和混合激励线性预测语音编码。

本书内容丰富，系统深入，原理和技术翔实，文献齐全，可作为高等院校通信与信息处理及相关专业的高年级本科生、硕士生和博士生的教材和参考书，也可供从事相关专业工作的工程技术人员和科研人员参考。

图书在版编目(CIP)数据

数字语音编码原理/鲍长春编著. —西安: 西安电子科技大学出版社, 2007. 1

(现代通信理论与技术丛书)

ISBN 978-7-5606-1760-2

I. 数… II. 鲍… 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 004640 号

策 划 臧延新

责任编辑 邵汉平

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

http: //www. xduph. com E-mail: xdupfxb@pub. xaonline. com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2007 年 1 月第 1 版 2007 年 1 月第 1 次印刷

开 本 787 毫米×1092 毫米 1/16 印张 21

字 数 496 千字

印 数 1~4000 册

定 价 26.00 元

ISBN 978-7-5606-1760-2/TN·0354

XDUP 2052001-1

*** 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

语音编码在数字通信网中占有重要的位置,人们对移动电话和 IP 电话持续增长的需求有力地证明了它的生命力。随着信息社会和通信技术的高速发展,频率资源愈发显得宝贵。因此,压缩语音信号的传输带宽或降低电话信道的传输码率,一直是人们追求的目标,语音编码在实现这一目标中担当着重要角色。

本书是作者以 2001 年 3 月出版的《低比特率数字语音编码基础》(北京工业大学出版社出版)一书为基础,经过补充和修改多个章节而成的。全书系统而深入地介绍了低比特率数字语音编码的基本理论和基本原理,内容丰富,系统深入,原理和技术翔实,文献齐全。

全书共 12 章。第 1 章为语音编码概论;第 2 章为语音信号的数字模型;第 3 章为语音信号的数字分析;第 4 章为语音信号的基音检测;第 5 章为语音信号的线性预测分析;第 6 章为语音信号的矢量量化方法;第 7 章为线性预测系数的有效量化方法;第 8 章为码激励线性预测语音编码;第 9 章为波形内插语音编码;第 10 章为多带激励语音编码;第 11 章为正弦语音编码;第 12 章为混合激励线性预测语音编码。其中第 2~7 章为低比特率数字语音编码的基本理论,第 8~12 章为经典的低比特率数字语音编码原理和技术。

本书的特点是以语音信号产生的源-系统模型为“绳”,先介绍语音特征参数的分析、提取和量化方法,再根据激励源的不同,系统而深入地介绍国际上热衷研究并得到应用的五种语音编码技术。通过本书的学习,读者可以掌握低比特率语音编码的基本知识和理论,了解和掌握当前语音编码的前沿技术。

另外,将有关语音信号处理技术的综述、本书的第 2~7 章以及第 8~12 章中的任意一章组合,可作为高年级本科生或硕士研究生 32~40 学时的“语音信号处理”课程的教材(作者从 1998 年开始尝试这种安排,并安排了 5 个典型的上机大作业,收到了良好的教学效果);将第 8~12 章与有关文献组合,可作为博士生 40~60 学时的“现代语音编码技术”前沿课程的教材或参考书。

本书的出版得到了西安电子科技大学出版社的大力支持,在此表示衷心的感谢。本书引用了大量文献资料,在此也向原作者表示深深的谢意。

由于作者水平所限,疏漏和错误之处在所难免,敬请读者批评指正。

作 者

2006 年 11 月于北京工业大学

目 录

第 1 章 语音编码概论	1
1.1 引言	1
1.2 语音编码器的基本属性	2
1.2.1 语音质量	2
1.2.2 编码速率	4
1.2.3 编解码的复杂度	4
1.2.4 编解码延时	4
1.3 标准化的语音编码器	5
1.3.1 ITU-T 语音编码标准	5
1.3.2 欧洲数字蜂窝电话标准	6
1.3.3 北美数字蜂窝电话标准	7
1.3.4 日本数字蜂窝电话标准	8
1.3.5 保密通信电话标准	8
1.3.6 卫星电话标准	9
1.3.7 语音编码器的性能比较	9
1.4 语音编码的发展方向	11
第 2 章 语音信号的数字模型	13
2.1 发音器官	13
2.1.1 肺和气管	13
2.1.2 喉	14
2.1.3 声道	15
2.2 语音产生过程	16
2.3 语音的声学特性	17
2.4 语音信号产生的数字模型	18
2.5 听觉器官和语音感知	19
2.5.1 听觉器官	19
2.5.2 语音感知	20
第 3 章 语音信号的数字分析	22
3.1 语音信号的数字化	22
3.1.1 取样和量化	22
3.1.2 语音信号的预处理及时域波形	25
3.2 短时平均能量、振幅和跨零数	29
3.2.1 短时平均能量	29
3.2.2 短时平均幅度函数	31
3.2.3 短时平均跨零数(过零率)	32
3.3 短时自相关函数	33
3.4 短时傅里叶变换(STFT)	35
3.4.1 傅里叶变换解释	35

3.4.2 滤波器解释	41
3.5 短时傅里叶变换的取样率	43
3.5.1 时域取样率	43
3.5.2 频域取样率	44
3.5.3 时域和频域总取样率	45
3.6 语音信号的短时综合	45
3.6.1 滤波器组相加法	45
3.6.2 叠接相加法	49
3.7 语音信号的语谱图分析	52
3.8 语音信号的倒谱分析	54
第4章 语音信号的基音检测	58
4.1 基音检测的现状	58
4.1.1 时域波形类似性方法	59
4.1.2 频域谱类似性方法	59
4.2 基音检测的预处理和后处理	60
4.2.1 谱平整	60
4.2.2 基音轨迹平滑	61
4.3 基于归一化互相关函数的基音检测算法	62
4.3.1 归一化互相关函数的确定	62
4.3.2 预处理	63
4.3.3 后处理	70
4.3.4 清浊判决	70
4.3.5 实验结果	71
4.4 基于 MBE 模型的基音检测算法	73
4.4.1 频域估计	73
4.4.2 时域近似估计	75
4.4.3 偏差校正	76
4.4.4 算法的实现	79
4.5 基于正弦模型的基音检测算法	86
4.5.1 算法的基本原理	86
4.5.2 改善性能的几种措施	89
4.5.3 短时谱包络估计	90
4.5.4 浊音度(voicing)检测	90
4.5.5 实验结果	91
第5章 语音信号的线性预测分析	93
5.1 线性预测分析的基本原理	93
5.1.1 LPA 的实现方法	93
5.1.2 语音信号模型参数与线性预测误差滤波器参数之间的关系	95
5.2 线性预测分析的解法	97
5.2.1 自相关法	98
5.2.2 协方差法	101
5.2.3 格型法(或协格法)	106
5.3 线性预测分析的应用	109

5.3.1 LPA 谱估计	109
5.3.2 语音信号 LPA 谱估计的实际考虑	111
5.3.3 基音检测	113
5.4 线谱频率分析	116
5.4.1 线谱频率参数的定义及性质	116
5.4.2 LSF 参数与 LPA 系数间的相互转换	122
第 6 章 语音信号的矢量量化方法	128
6.1 矢量量化的基本原理	128
6.1.1 矢量量化的定义与矢量量化器的构造	128
6.1.2 最佳矢量量化器	131
6.1.3 矢量量化器的设计算法	132
6.2 矢量量化系统	135
6.2.1 全搜索矢量量化器	135
6.2.2 树搜索矢量量化器	137
6.2.3 多级矢量量化器	139
6.2.4 乘积码矢量量化器	142
6.3 有记忆的矢量量化系统	147
6.3.1 预测矢量量化器	147
6.3.2 有限状态矢量量化器	151
第 7 章 线性预测系数的有效量化方法	154
7.1 失真测度及客观评价	154
7.1.1 失真测度	154
7.1.2 客观评价指标	156
7.2 分裂式矢量量化	157
7.3 多级矢量量化	158
7.4 预测分裂式矢量量化	161
7.5 连接分裂式矢量量化	164
7.6 去除非典型胞腔	169
第 8 章 码激励线性预测语音编码	171
8.1 码激励线性预测语音编码模型	171
8.2 激励参数的优化原理	175
8.2.1 两级码书结构的 CELP 编码器	175
8.2.2 激励码书的联合优化和分级优化	177
8.2.3 激励参数优化的基本计算量	178
8.3 固定码书搜索算法	179
8.3.1 特殊设计的码书搜索算法	179
8.3.2 自相关算法	180
8.3.3 变换域算法	181
8.3.4 多级搜索算法	183
8.3.5 码字叠接递推算法	183
8.3.6 代数码书搜索算法	187
8.3.7 矢量和码书搜索算法	189

8.4 自适应码书搜索算法	192
8.4.1 整数延时自适应码书的搜索	192
8.4.2 分数延时自适应码书的搜索	194
8.5 激励码书增益的量化	200
8.5.1 增益的归一化矢量量化	201
8.5.2 增益的预测式矢量量化	202
8.6 自适应后滤波	203
8.6.1 短时后滤波器	203
8.6.2 长时后滤波器	205
8.6.3 组合的后滤波器	209
8.6.4 后滤波器的性能	210
8.7 CELP 语音编码器实例	211
8.7.1 FS1016 4.8 kb/s CELP	212
8.7.2 G.728 16 kb/s LD-CELP	213
8.7.3 IS54 8 kb/s VSELP	214
8.7.4 JDC 3.6 kb/s PSI-CELP	216
8.7.5 G.729 8 kb/s CS-ACELP	217
第9章 波形内插语音编码	220
9.1 原型波形内插的基本思想	220
9.2 原型波形内插原理	221
9.2.1 瞬时波形及其内插	221
9.2.2 PCW 波形动态的自适应	223
9.3 原型波形内插的实际实现	225
9.3.1 原型波形谱包络的修正	225
9.3.2 类似性测量	226
9.3.3 对齐(Alignment)和 SCR	227
9.3.4 原型波形的量化	228
9.3.5 PW 的内插	230
9.3.6 PWI 和其它编码技术的过渡	231
9.4 原型波形的提取方法	232
9.4.1 在语音域中提取 PW	232
9.4.2 在残差域中提取 PW	233
9.5 PWI 编码器性能分析	233
9.6 特征波形内插的基本原理	235
9.7 特征波形内插的分析模块	236
9.7.1 线性预测分析(LPA)	237
9.7.2 基音估计	237
9.7.3 基音内插	238
9.7.4 特征波形的表达	239
9.7.5 特征波形的提取	241
9.7.6 特征波形的对齐	242
9.7.7 特征波形的功率计算及归一化	248
9.8 特征波形内插的合成模块	251

9.8.1	瞬时基音周期和瞬时 CW 的生成	251
9.8.2	相位轨迹估计	254
9.8.3	分析合成层的性能	254
9.9	CWI 编码器参数的量化	255
9.9.1	特征波形功率的量化	256
9.9.2	特征波形的量化	257
9.9.3	CWI 编码器的性能	262
第 10 章	多带激励语音编码	263
10.1	MBE 语音编、解码模型	263
10.2	MBE 的语音分析	266
10.2.1	基频估计	266
10.2.2	清/浊判决	267
10.2.3	谐波幅度估计	269
10.3	MBE 的语音合成	269
10.3.1	清音语音合成	270
10.3.2	浊音语音合成	270
10.3.3	重建语音的产生	272
10.4	MBE 模型参数的量化	272
10.5	INMARSAT-M 系统中的 4.15 kb/s IMBE 语音编码器	273
10.6	低比特率 MBE 语音编码器	275
10.6.1	MBE 幅度的 LPA 谱逼近	275
10.6.2	一种基于 LPA 的 2.4 kb/s MBE 声码器	279
第 11 章	正弦语音编码	283
11.1	广义正弦分析与合成模型	283
11.2	STC 语音编码	284
11.2.1	低速率 STC 语音编、解码原理	284
11.2.2	2.4/4.8 kb/s STC 语音编码比特分配及性能比较	287
11.3	HELP 语音编码	290
11.3.1	HELP 语音编、解码原理	290
11.3.2	HELP 语音编码关键技术	292
11.3.3	2.3 kb/s HELP 语音编码比特分配及性能比较	294
第 12 章	混合激励线性预测语音编码	296
12.1	LPC 语音合成模型	296
12.2	MELP 语音合成模型及机理	297
12.2.1	混合的脉冲和噪声激励	297
12.2.2	周期或非周期脉冲	300
12.2.3	自适应谱增强	301
12.2.4	脉冲散布滤波器	303
12.2.5	残差谐波幅度模型	303
12.3	MELP 编码器	305
12.3.1	参数分析与提取	305
12.3.2	参数量化与编码	309

12.4 MELP 解码器 311

12.4.1 信道解码 312

12.4.2 增益抑制 312

12.4.3 参数插值 313

12.4.4 混合激励信号的生成 313

12.4.5 自适应谱增强 314

12.4.6 线性预测合成 314

12.4.7 增益调整 314

12.4.8 脉冲散布滤波 314

12.4.9 合成部分的循环控制 315

12.5 2.4 kb/s MELP 语音编码的比特分配与性能比较 315

参考文献 316

第 1 章 语音编码概论

1.1 引 言

随着信息社会和通信技术的高速发展,频率资源愈发显得宝贵。因此,压缩数字语音信号的传输带宽或降低电话信道的传输码率,一直是人们追求的目标,语音编码在实现这一目标中担当着重要角色。语音编码就是压缩语音信号的数字表示而使表达这些信号的比特需求最小的算法^[1]。

语音编码目前主要分为三类,即波形编码、参数编码和混合编码。波形编码力图使重建语音波形保持原语音信号的波形形状,它通常将语音信号作为一般的波形信号来处理,具有适应能力强、话音质量高等优点,但所需要的比特率高,通常能在 16~64 kb/s 的速率上给出高的编码质量,当速率进一步降低时,其性能会迅速下降。参数编码则通过对语音信号特征参数的提取和编码,力图使重建语音信号具有尽可能高的可懂性,即保持原语音的语意,而重建语音信号的波形同原始语音信号的波形可能会有相当大的差别,但其编码速率很低,可低至 1.2~2.4 kb/s。参数编码的主要问题是合成语音质量差,特别是自然度较低;另外,它对讲话环境的噪声比较敏感,只有在安静的讲话环境下才能给出高的可懂度。混合编码则克服了波形编码和参数编码的弱点,同时又结合了它们各自的长处,在 4~16 kb/s 速率上能够得到高质量的合成语音,在本质上它也具有波形编码的优点。

尽管通信网络的容量在不断增加,但语音压缩编码一直是现在和未来应用中令人感兴趣的,这是因为它有着广泛与强烈的应用背景。这些应用体现为如下两类:

(1) 语音信号的数字传输。该类应用主要有数字通信系统(Digital Communications Systems)、移动无线电(Mobile Radio)、蜂窝电话(Cellular Telephony)和保密话音系统(Secure Voice Systems)。这类应用又称为数字电话通信系统或数字语音通信系统,与模拟语音通信系统相比数字语音通信系统具有抗干扰强,保密性好,易于集成化等优点。它要求能够实时编解码,有高的抗信道误码能力,能传输带内数据、单频和多频等非语音信号,并具有多次音频转接的能力。信道条件、时延和数据速率是在该类应用中重要的考虑对象。图 1.1 为该类应用的系统框图。

(2) 语音信号的数字存储。该类应用主要有呼叫服务(Paging Services)、数字应答机(Digital Answering Machines)和声音响应系统(Voice Response Systems),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类电子发声玩具等。这类应用又称为数字语音录放系统,它与模拟语音录放系统相比具有灵活性高,可控性强和寿命长等优点。这类应用中对编码器的实时性要求不高,即不一定要实时编码,但希望有较

高的压缩效率，以降低所需存储器的容量。对解码器而言，则要求算法尽量简单，成本要低，并能够实时解码。在该类应用中人们最关心的是语音质量和存储需求。图 1.2 给出了该类应用的系统框图。

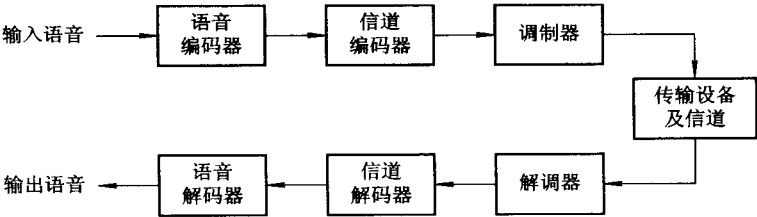


图 1.1 数字语音通信系统框图

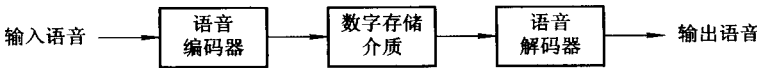


图 1.2 数字语音录放系统框图

近十几年来，语音编码取得了突飞猛进的发展，在国际标准化工作中堪称最为活跃的领域，已具备比较完善的理论和技术体系，并进入实用阶段。本章接下来简要介绍语音编码器的属性、现存的一些语音编码标准和语音编码的发展方向。

1.2 语音编码器的基本属性^[2]

语音编码器是基于某种应用而开发的，不同的应用场合将使用不同的编码器。在优化不同应用场合的语音编码器时，需要联动考虑语音编码器的四个基本属性，即语音质量、比特率、复杂度和延时。语音编码研究的基本问题就是在给定编码速率的条件下，如何得到尽可能好的重建语音质量，并保证尽可能小的编解码延时和适当的算法复杂度。或者说是给定编码质量、编解码延时及算法复杂度的条件下，如何降低语音编码所需的比特率。这四个因素之间有着密切的联系，并且在不同的应用中对各方面的侧重要求也有所不同。

1.2.1 语音质量

重建语音信号的质量是语音编码器至关重要的一个属性。语音质量的评价是一个非常困难的问题，至今还没有找到一种客观的评价方法能很好地将各种语音编码器的质量和输入语音信号联系起来。也就是说，随着比特率的降低，编码器重建信号的质量变得越来越依赖于输入信号的特征，这样就使得预测一个将应用于实际场合的语音编码器的质量变得非常困难。所以，在评价某个语音编码器的实际适用性时，需要进行广泛的人为主观评价。

语音质量主观评价主要采用统计的方法进行，对高质量的语音通常要考虑语音的清晰度和自然度等。清晰度是衡量话音中的字、单词和句子的可懂度的参量，自然度则是对讲话人的辨识水平的反映。主观评价方法有平均意见分测试(Mean Opinion Score, MOS)、判断韵字测试(Diagnostic Rhymer Test, DRT)和判断满意度测试(Diagnostic Acceptability Measure, DAM)等。

平均意见分测试(MOS)通常采用 5 级评分标准,即优(Excellent)、良(Good)、中(Fair)、差(Poor)和坏(bad),可用数字 5~1 表示这 5 个等级。其中:5 对应于优,这时失真察觉不到;4 对应于良,这时失真刚有察觉;3 对应于中,这时失真有察觉且稍觉可厌;2 对应于差,这时失真能明显察觉到,且可厌,但可忍受;1 对应于坏,这时失真不可忍受。参加测试的人员在听完所测语音后,从这 5 个等级中选择其中某一级作为他对所测语音的评价。全体参测人员的平均分就是所测语音质量的 MOS 分。

采用 MOS 打分通常有三个优点:

- (1) 可以评定各种各样的语音损伤;
- (2) 能直接反映用户的意见;
- (3) 不必要求特别受训人参加打分,打分容易进行。

然而打分也有一些缺点,即打分易受测试条件的影响,如打分者的选择,测试语句的选择,对参加测试人的指导,演示时语句的先后以及演示的方法等;而且,对于不同国籍的人,对其语音质量的打分也有所不同。由于目前没有代替它的更好方法,因此,它仍然是一种主要的主观语音质量评价方法。

在数字语音通信中,通常认为,MOS 分在 4.0~4.5 时为高质量数字化语音,达到长途电话网的质量要求,接近于透明信道编码,也常称之为网络质量。MOS 分在 3.5 左右时称做通信质量,这时能感到重建语音质量有所下降,但不妨碍正常通话,可以满足多数话音通信系统的使用要求。MOS 分在 3.0 以下时,称为合成语音质量,这时重建语音具有足够的可懂度,但自然度及讲话人的确认等方面不够好。表 1.1 概括了 MOS 评价等级。

表 1.1 MOS 评价等级

MOS 级	主 观 意 见	质 量
5(优)	失真察觉不到	透明
4(良)	失真刚有察觉	长途
3(中)	失真有察觉且稍觉可厌	通信
2(差)	失真能明显察觉到且可厌,但可忍受	合成
1(坏)	失真不可忍受,非常可厌	坏

判断韵字测试(DRT)是反映话音清晰度或可懂度的一种方法,它主要应用于低速率语音编码的质量测试,因为这时的可懂度已成为主要问题。这种测试方法使用若干对同韵母进行测试,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母中的某个音,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 得分为 95%以上时,语音清晰度为优,85%~94%为良,75%~84%为中,65%~74%为差,而 65%以下为不可接受。在实际通话中,清晰度为 50%时,整句的可懂度为 80%,这是因为整句中具有很高的冗余度,即使个别字听不清楚,人们也能够理解整句话的意思。当清晰度为 90%时,整句话的可懂度接近 100%,所以,对于低速率语音编码,一般要求其清晰度能够达到 90%以上。

判断满意度测试(DAM)是对语音质量的综合评估,它是在许多条件下对语音质量可接受程度的一种量度,也采用百分比评分。

1.2.2 编码速率

编码速率可以用“比特/秒”(b/s 或 bps)来度量,它代表了编码的总速率,一般用 I 表示。编码速率也可以用“比特/样点”(b/p)来度量,它代表了平均每个语音样点用多少个比特编码,一般用 R 表示。 I 和 R 可以通过采样率联系起来:

$$I = R \cdot f_s$$

其中,采样率 f_s 是根据采样定理由信号带宽决定的。目前国际上数字音频信号常用的带宽及采样率如表 1.2 所示。

表 1.2 数字音频带宽及采样率

信号类型	采样率/kHz	带宽/kHz	频率范围/Hz
电话	8	3.2	200~3400
电话会议	16	7	50~7000
激光唱盘(CD)	44.1	20	20~20 000
数字语音广播及数字语音磁带	48	20	20~20 000

很明显,平均每样点比特数 R 越高,语音波形或参数量化则越精细,语音质量也就越高,相应地,对传输带宽或存储容量的要求也就越高。在目前的波形编码中,为了获得高质量的重建语音, R 一般要大于 2,但在参数编码中, R 可以是几分之一比特或更低,这时的重建语音质量在自然度方面会有很大的损失。

1.2.3 编解码的复杂度

编解码算法的复杂程度与语音编码的语音质量有着密切的联系。在相同比特率的情况下,采用相对复杂的一些算法会获得更好的语音质量,或者说,对于相同的语音质量,采用相对复杂的算法会降低编码比特率。

编解码算法的复杂度与硬件的实现有着密切的关系,它决定了硬件实现的复杂程度、体积、功耗及成本。目前的很多语音编码算法都采用通用 DSP 芯片来实时实现,这主要是考虑了投资少,研制周期短等优点。语音编码算法的复杂度对 DSP 芯片的运算能力及存储器容量(RAM)提出了具体的要求,语音编码算法的复杂度越高,则所需 DSP 芯片的速度越快,存储器容量越大,成本和功耗也越高。DSP 芯片的运算能力一般可用每秒执行的百万条指令(MIPS)来衡量,存储器容量通常用千字(K Word)或千字节(KB)的数量来估计。如果大量使用,则应考虑开发和使用 VLSI 器件或专用芯片。

1.2.4 编解码延时

在数字语音通信系统中,语音编解码延时像线路传输延时一样,对系统的通话质量有很大的影响。一般地,编解码的算法越复杂,延时越大。延时过大,会明显感到通话对方反应“迟钝”,甚至导致正常通信困难。延时影响通话质量的另一个原因是回声。在固定电话网的二/四线转换中,由于阻抗匹配不理想,发送端收到的语音信号会泄露到该端四线的发送路径中,在接收端就形成了回声。当延时比较小时,回声同话机侧音及房间交混回响

声相混,因而感觉不到;当往返总延时超过 100 ms 时,发话者就能从手机中听到自己的回声。如果回声传输路径衰减不够大,则就会听到多次回声,从而严重影响通话质量。

对于公众电话网,可能会有几次音频转接,即会有多次语音编解码,因此,对于单次语音编解码延时,通常要求不超过 5~10 ms。对于其它一些类型的通信系统,如卫星通信,由于传输延时或纠错编码引入的延时等已经比较大,因此再苛求语音编码的延时就意义不太大了,这时主要强调语音编码质量、编码速率及抗误码性能等要求,而允许延时可为几十毫秒到 100 毫秒。为了保证正常通信,当总延时超过 100 ms 时,应采取回声抵消或回声抑制等措施。

语音编码器的一路编解码延时通常由独立的四部分组成^[3],即算法延时(编码器缓冲延时)、计算延时(编码器处理延时)、传输延时(译码器缓冲延时)和译码器处理延时。算法延时等于语音信号处理前所采集的语音信号的长度,即语音信号的分析帧长;计算延时等于编码器处理分析帧数据而形成比特流的时间,该延时必须小于编码器缓冲延时,否则编码器不能处理下一帧的数据;传输延时等于译码器为进行解码而等待接收比特流的时间;译码器处理延时等于译码比特流而得到合成语音的时间,该延时不能大于编码器缓冲延时,否则不能得到下一帧的合成语音。

1.3 标准化的语音编码器^[1, 4]

人们对通信方式共同的强烈需要是语音编码标准存在的原因。制造商、服务商和消费者都认识到,他们共同感兴趣的是有一个公共标准。对于一个假想的新应用来说,不同的卖主对同样的目的可以建议非常不同于他人的私有标准,这样将导致不同卖主的产品相互不兼容,难以实现一个私有标准完全占领市场的目的。只有在标准框架上达成协议,形成新的标准,才能真正实现不同卖主的利益共享。

目前,从事语音编码标准制定的主要组织有国际电信联盟电信标准部(International Telecommunications Union - Telecommunication Standardizations Sector, ITU - T)、欧洲电信标准协会(European Telecommunications Standards Institute, ETSI)、北美电信工业协会(Telecommunications Industry Association, TIA)、日本的无线电系统研究和开发中心(Research and Development Center for Radio System, RCR)、管理全世界地球同步通信卫星的国际海事卫星公司(International Maritime Satellite Corporation, INMARSAT)、制定保密电话语音编码标准的美国政府和北大西洋公约组织(North Atlantic Treaty Organization, NATO)等。

1.3.1 ITU - T 语音编码标准

传统上 ITU - T 主要是为 8 kHz 采样率和 3.4 kHz 带宽的 PSTN 电话制定的语音编码标准,目的是借助数字电路的多路复用技术来提高通信网的容量。现在,ITU - T 已经在制定 16 kHz 采样率和 7 kHz 带宽的语音编码标准,目的是实现 ISDN、无线移动通信网和 IP 网的互通和兼容。

1972 年,ITU - T 发布了 A/ μ 律 64 kb/s PMC(Pulse - Code Modulation)语音编码标准 G. 711^[5]。该标准的基本思想是,先将 8 kHz 采样率的离散语音样点变换到对数域,然

后对变换的语音样点进行 8 bit 量化编码。作为第一个数字电话系统, G. 711 已经配置于世界上不同的 PSTN 中。从那时起, ITU - T 开始制定更复杂的 G. 72x 系列语音编码标准, 颁布了 G. 721 32 kb/s ADPCM (Adaptive Differential Pulse Code Modulation) 编码标准及其扩展标准 G. 726 (40/32/24/16 kb/s)^[6], 并最终用 G. 726 取代了 G. 721。除了 G. 723. 1^[7], ITU - T 的每个语音编码器都是以前一个标准的半速率为基础进行开发的, 例如, 1992 年和 1996 年颁布的 G. 728^[8] 和 G. 729^[9] 语音编码器的速率分别为 16 kb/s 和 8 kb/s。另外, ITU - T 还颁布了用于可视电话的 5. 3/6. 3 kb/s 双速率语音编码标准 G. 723. 1。G. 728、G. 729 和 G. 723. 1 的编码原理均基于码激励线性预测 (Code Excited Linear Prediction, CELP) 技术。为了支持非连续传输, ITU - T 还分别公布了 G. 729 和 G. 723. 1 的扩展标准 G. 729B^[10] 和 G. 723. 1A^[11]。基于它们的静默压缩方案, 这些标准已广泛应用于包交换的语音通信中^[12]。在过去的七、八年中, ITU - T 一直进行着 4 kb/s 语音编码标准的制定工作, 目前已有两种编码器成为候选标准, 一种编码器完全基于 CELP 模型^[13], 而另一种编码器模型则是一种基于 CELP 和正弦语音编码原理^[14, 15] 的混合模型。由于这两个编码器还没有满足 ITU - T 所列参考项的全部要求, 因此 4 kb/s 语音编码标准的颁布还有待时日。表 1. 3 概括了 ITU - T 窄带语音编码标准。

表 1. 3 ITU - T 窄带语音编码标准

编码标准	G. 711 A/ μ 律 PCM	G. 726 ADPCM	G. 728 LD-CELP	G. 729 CS-ACELP	G. 723. 1 ACELP/MP-MLQ	G. 4K (待定)
公布年代/年	1972	1990	1992	1996	1995	
比特率/(kb/s)	64	40/32/24/16	16	8	5. 3/6. 3	4
话音激活检测 (VAD)	无	无	无	有	有	未知
话音质量	长途	长途	长途	长途	长途	长途
算法延时/ms	0. 125	0. 125	0. 625	15	37. 5	25

上述标准针对的是 300~3400 Hz 的语音带宽, 此时的通话质量尚不足以满足召开音频会议时所希望达到的那种面对面交谈的语音质量。所以, ITU - T 又制定了两个宽带语音编码标准 G. 722^[16] 和 G. 722. 1^[17], 主要用于高质量的多媒体通信。G. 722 基于窄带 ADPCM (SB-ADPCM) 定义了 64 kb/s、56 kb/s、48 kb/s 三种比特率, 它利用正交镜像滤波器将输入信号分解为低子带和高子带, 然后根据不同的子带, 借助变步长的 ADPCM 方法量化带通滤波的语音信号; 而 G. 722. 1 则基于变换编码技术定义了 32 kb/s 和 24 kb/s 两种速率。目前, ITU - T 正着手制定具有 5 层速率 (8 kb/s、12 kb/s、16 kb/s、24kb/s 和 32 kb/s) 的嵌入式变比特率宽带语音编码标准^[18]。

1. 3. 2 欧洲数字蜂窝电话标准

随着数字蜂窝电话的出现, 欧洲电信标准协会 (ETSI) 制定了许多语音编码标准。

ETSI 相继公布了 13 kb/s GSM 全速率(Full Rate, FR)语音编码标准^[19]、5.6 kb/s GSM 半速率(Half Rate, HR)语音编码标准和 12.2 kb/s GSM 增强的全速率(Enhanced Full Rate, EFR)语音编码标准^[20, 21]。继这三种标准之后,ETSI 又公布了一种自适应多速率(Adaptive Multi-Rate, AMR)语音编码标准^[22],该标准共有 8 种速率(从 12.2 kb/s 到 4.75 kb/s),其中 4 个速率用于全速率信道,而另外 4 个速率用于半速率信道。AMR 编码器的目的是根据信源编码和信道编码间的最佳选择提供增强的语音质量,在高的无线干扰时,为保障语音质量,AMR 能够自动减少信源编码比特数,并将富裕的比特数补充到信道编码中,反之亦然。

ETSI 语音编码器也能够借助于语音激活检测^[23-26]进行静默压缩,这非常有利于减少移动通信中的信道干扰和延长电池寿命。表 1.4 概括了欧洲移动通信中使用的标准语音编码器。

表 1.4 用于 GSM 移动通信的 ETSI 语音编码标准

语音编码器	全速率(FR) RPE-LTP	半速率(HR) VSELP	增强全速率(EFR) ACELP	自适应多速率(AMR) ACELP
公布年代/年	1987	1994	1998	1999
比特率/(kb/s)	13	5.6	12.2	12.2/10.2/7.95/7.4/ 6.7/5.9/5.15/4.75
语音激活检测(VAD)	有	有	有	有
语音质量	接近长途	接近长途	长途	通信
算法延时/ms	20	24.375	20	25

1.3.3 北美数字蜂窝电话标准

在北美,电子工业协会(Electronic Industries Association, EIA)下属的北美电信工业协会(TIA)根据美国使用的码分多址(Code Division Multiple Access, CDMA)和时分多址(Time Division Multiple Access, TDMA)技术对其移动通信进行了标准化。1993 年,TIA/EIA 采纳了 Qualcomm CELP(QCELP)^[27]作为过渡语音编码标准 IS-96-A(Interim Standard-96-A),该标准对语音信号进行 8 kb/s 到 0.8 kb/s 的变速率编码。1995 年,TIA/EIA 在 IS-96-A 的基础上颁布了一个称为 IS-127^[28]的增强型变速率(Enhanced Variable Rate Coder, EVRC)语音编码标准,该标准在语音压缩模块的前端引入了一个新颖的语音增强功能,在噪声背景条件下,获得了更加舒适的语音质量。1998 年,TIA/EIA 公布了用于个人通信系统的变速率语音编码标准 IS-733^[29],该标准的速率介于 14.4 kb/s 和 1.8 kb/s 之间。1989 年和 1996 年,TIA/EIA 又分别颁布了用于北美 TDMA 的 7.95 kb/s 全速率语音编码标准 IS-54^[30]和增强的全速率语音编码标准 IS-641-A^[31]。表 1.5 概括了北美移动通信中使用的标准语音编码器。

表 1.5 用于北美 CDMA/TDMA 移动通信的 TIA/EIA 语音编码标准

语音编码器	IS-96-A QCELP	IS-127 EVRC	IS-733 QCELP	IS-54 VSELP	IS-641-A ACELP
公布年代/年	1993	1995	1998	1989	1996
比特率/(kb/s)	8.5/4/2/0.8	8.5/4/2/0.8	14.4/7.2/3.6/1.8	7.95	7.4
语音激活检测(VAD)	有	有	有	有	有
语音质量	接近长途	长途	长途	接近长途	长途
算法延时/ms	25	25	25	25	25

1.3.4 日本数字蜂窝电话标准

日本数字蜂窝(Japanese Digital Cellular, JDC)电话标准由日本的无线电系统研究和开发中心(RCR)进行标准化,作为日本的 TDMA 数字移动电话语音编码标准。1990 年,RCR 公布了一种类似于 IS54 的 6.7 kb/s VSELP 编码器^[30]作为 JDC 的全速率语音编码标准。为了使日本 TDMA 个人数字移动系统(Personal Digital Cellular, PDC)的容量增加一倍,1993 年 RCR 又公布了 JDC 3.6 kb/s 基音同步更新码激励线性预测(Pitch Synchronous Innovation-Code Excited Linear Prediction, PSI-CELP)编码器^[32, 33]作为 JDC 的半速率语音编码标准。表 1.6 概括了 JDC 语音编码标准。

表 1.6 JDC 语音编码标准

语音编码器	JDC(VSELP)	JDC(PSI-CELP)
公布年代/年	1990	1993
比特率/(kb/s)	6.7	3.6
语音激活检测(VAD)	无	无
语音质量	<IS54	等同于全速率
算法延时/ms	25	50

1.3.5 保密通信电话标准

语音编码是保密通信系统中至关重要的一部分,在突发事件中,为了传达准确的语音命令,语音编码器的可懂度是一个主要的考虑对象。目前,保密通信的语音编码标准主要由美国国防部(Department of Defense, DoD)组织制定。1984 年和 1991 年,DoD 分别颁布了称之为 2.4 kb/s LPC-10e 编码器^[34-38]和 4.8 kb/s CELP 编码器^[39-42]的联邦标准 FS-1015(Federal Standard 1015)和 FS-1016。DoD 最新的保密电话标准^[43, 44]是基于正弦语音编码模型的 2.4 kb/s 混合激励线性预测(Mixed Excitation Linear Prediction, MELP)声码器(Vocoder)^[45-49],2.4 kb/s DoD MELP 语音编码器在半速率情况下给出了好于 4.8 kb/s FS-1016 编码器的语音质量。一个修正和改进的 2.4/1.2 kb/s MELP 编码器已经成为 NATO 的保密电话标准,该标准在编码前端使用了一个噪声预处理器。由于 MELP 编码器能在非常低的比特率产生可懂的语音质量,因此这类参数编码器已广泛使用在保密通信系统中。表 1.7 概括了 DoD 的标准语音编码器。