

新世纪信息与通信技术教程

# 数字语音—语音编码实用教程

李昌立 吴善培 编著

人民邮电出版社



# 前摇摇言

数字语音通信是当前信息产业中发展最快、普及面最广的业务。随电话的持续发展,使国内长途电话和国际长途电话大幅降价,广大人民受益。移动通信数字化,扩展了信道容量,改善了服务质量,促进了这一产业的飞速发展。在卫星信道、短波信道和有线信道,语音信号的数字化和压缩对军事通信更有其重要意义。总之,数字语音通信和模拟语音通信相比,具有抗干扰性强、保密性好、易于开展增值业务、管理和控制方便、生产成本低等优点。而语音的存储技术和语音编码又密不可分。语音应答系统、数字录音电话、语音信箱、语音查询系统等,也都是随着语音的压缩编码而发展起来的。

随着我国经济的发展,信息时代已经到来。人们对通信的要求越来越高,因此,频率资源也就愈加宝贵。在保证语音质量的前提下,降低传输的数码率,就意味着降低成本。因此,语音的高效编码,仍将是广大通信工作者长期奋斗的目标。

近十年来,由于数字信号处理算法和器件(集成电路)的飞速发展,为语音编码器的实现和应用奠定了基础。又由于通信对于标准化的迫切要求,语音编码的标准化取得了很大进展。国际电联的语音编码技术已经标准化,进入了实用阶段。国际电联的语音编码器,已有一些地区性标准,仍将是制定国际标准时竞争的热点。而今后的研究方向将会逐步转向更低的码率。

本书的目标是要让读者深入了解语音编码国际标准中的一些关键算法和技术,以及掌握这些技术所必须的基础知识、物理概念、计算方法和实验结果。只有掌握了一些基本的和成熟的方法,才有可能开发新的方案和算法,本书也介绍当前正在发展的一些新算法,以便为我国进一步和国际接轨,开发具有自主知识产权的产品,提供牢固坚实的知识基础。

语音编码是一门实践性很强的学科,不但要有方案和算法的理论研究,大量工作是在计算机上,对不同方案进行模拟、比较和改进,通常还要设计专门的实验平台,并对男女甚至儿童多个发音人的大量语音材料进行试验。国内外研究和开发的经验表明,提高语音编码技术的关键之一,就是要建立语音质量评价标准和严格的测试方法。本书将详细讨论这些标准和方法。

本书的另一个特点是由科研人员和教学人员共同编写,充分体现两者相结合的优点。即保持本书的系统性,让读者能够很好地掌握基本概念和理论。同时,对于实时编码器的软件硬件开发、系统集成、主客观评价试验、信道试验等,也能给读者提供一些方法和经验。

本书的内容主要来自国外有关语音编码的专著,国际电联对语音编码的建议,以及我们科研和教学的经验。最重要的参考书如下:

- \* 语音编码: 国际电联, 1988年, 北京人民邮电出版社, ISBN 7-113-01111-1
- \* 语音编码: 国际电联, 1988年, 北京人民邮电出版社, ISBN 7-113-01111-1
- \* 语音编码: 国际电联, 1988年, 北京人民邮电出版社, ISBN 7-113-01111-1



## 作者简介

李昌立 1956年毕业于北京邮电学院无线电工程系。中国科学院声学研究所研究员,曾任声学所语音和通信研究室主任。中国电子学会信号处理分会委员、中国通信学会信号信息处理委员会副主任。研究领域包括:语音信号处理、多媒体通信终端和系统、宽带移动通信工程中的开发和应用等。曾获国家发明创造二等奖、中科院科技一等奖、国家科学大会奖等。在《电子学报》、《声学学报》等刊物发表文章 100 多篇。

吴善培 1956年毕业于北京邮电学院有线电通信工程系,留校任教至今。1985~1986年在法国国立高等电信学校做访问学者,1989年又到法国做高级访问学者。现为北京邮电大学电信工程学院教授。一直从事通信与信息处理领域的教学与科研工作,主要研究语音编码与语音识别技术,在国内外技术会议及期刊上共发表学术论文 100 多篇。

# 序 摇摇言

语音编码最初是由人们企图压缩通信频带而来。自从 1876 年贝尔( Alexander Graham Bell) 发明电话以来, 电话通信发展很快, 但是在几乎半个世纪中, 人们对电话系统了解甚少, 在 20 世纪 40 年代, 美国贝尔实验室开始研究电话信号—语言和电话要求—听觉的特性, 同时也对电话系统的理论进行研究。不久即发现电话通信要求频带很宽, 只有 3000-8000 赫兹才能在电话上听懂, 但语音信号的变化没有那么快。人讲话时所有动作不过是声带( 声带) 的松紧、口腔的大小、软腭的开关、舌的前后、上下齿的接紧, 以及上下唇的开合等, 每一种的动作都不快, 只是每秒钟十次、二十次而已。所有动作加起来不到 1000 赫兹。为什么通信系统要求频带宽几十倍? 认为压缩频带潜力很大。香农( Claude Shannon) 等提出信息论, 并做出语音演示器( 语音合成器) 和声码器( 语音编码器) 到 1937-1940 年才开始公开演示。这就是语音编码的理论和实践的开端( 演示器和声码器的数字化还稍后)。

信息论和这些系统已有长足发展, 达到成熟的程度。压缩频带也有巨大发展, 例如有些实验达到一个字母用一位( 1/2 或半位)。但是更重要的是信号处理技术的普遍应用, 在现有高速微型计算机的条件下, 频带已不是基本问题。当前信号处理技术的成就使得几乎大半物理和工程上的研究问题都可以在微型计算机上完成。

语音编码是信号处理技术的一个方面, 已成为语音通信的主要技术。本书透彻地阐明语音编码技术的基础知识、物理概念、一些关键的算法和技术, 以及实验结果。这就使读者有了充分的准备, 熟练地正确使用语音编码技术和进一步发展这些技术。

从 20 世纪 70 年代起, 人们做了巨大努力去解决语言自动识别的问题, 虽经政府和企业的重大支持但几次集中努力都失败了。这证明语言自动识别不仅仅是技术问题, 全靠数字信号处理技术不可能解决。后来投入大量人力, 用了好多年研究“语言学”、汉语的特性和特点, 于 1983 年制造了“语音识别器”成为基本识别汉语的设备, 现已制成软件, 可用到微型计算机上, 这证明在语言信号处理上, “语言学”也可取得重要应用, 不知在语音编码方面如何, 还有待进一步研究。

中国科学院院士

马大猷  
2004 05 01

# 目 录

第 1 章 语音编码概论	1
1.1 语音编码的数字化	1
1.2 语音编码的必要性和推动力	1
1.3 语音编码的声学基础知识	1
1.3.1 响度级和响度	1
1.3.2 听觉掩蔽和临界频带	1
1.3.3 听觉掩蔽在语音编码中的作用	1
1.3.4 发声器官的声电类比线路	1
1.3.5 语言发声的数学模型	1
1.4 语音信号的物理特性	1
1.4.1 语音信号在时域和频域的特性	1
1.4.2 语音信号的统计特性	1
1.4.3 语音传输所需比特率的估计	1
1.5 除去语音冗余度的方法	1
1.6 语音编码器方案的分类	1
1.6.1 波形近似语音编码器	1
1.6.2 参量编码器	1
1.7 语音编码的发展趋势	1
1.8 参考文献	1
第 2 章 语音信号的采样和量化	2
2.1 采样	2
2.2 标量量化	2
2.2.1 量化误差	2
2.2.2 均匀量化器	2
2.2.3 最佳量化器	2
2.2.4 对数量化器	2
2.2.5 自适应量化器	2
2.2.6 差分量化器	2
2.3 矢量量化	2
2.3.1 失真测度	2
2.3.2 码书设计	2
2.3.3 码书种类	2
2.3.4 训练试验和码书的顽健性	2
2.4 参考文献	2

第 猿章 摇语音信号的分析和建模 .....	源
摇摇猿 摇短时频谱分析 .....	源
猿 摇窗的作用 .....	源
摇摇源 摇语音信号的线性预测模型 .....	缘
猿 摇语音产生的源滤波器模型 .....	缘
猿 摇线性预测分析(蕴云)的求解 .....	缘
猿 摇线性预测分析的实现和物理解释 .....	源
摇摇缘 摇音调预测 .....	源
猿 摇语音信号的周期性 .....	源
猿 摇音调预测器 .....	缘
摇摇源 摇音调检测 .....	源
猿 摇时域波形的相似性方法 .....	源
猿 摇频率域频谱的相似性方法 .....	源
猿 摇预处理和后处理技术 .....	源
摇摇参考文献 .....	源
第 源章 摇使用线谱频率(蕴云)量化线性预测编码(蕴悦)参量 .....	愿
摇摇源 摇蕴悦参量的编码 .....	愿
摇摇源 摇蕴悦参量的另外一种表示——孕粤粤谱系数 .....	愿
摇摇源 摇蕴悦到蕴云的变换 .....	愿
猿 摇蕴云的计算 .....	愿
猿 摇蕴云到蕴悦的变换 .....	怨
猿 摇蕴云的特性 .....	怨
摇摇源 摇转换型——自适应帧间矢量预测(杂孕) .....	愿
猿 摇杂孕编码方案 .....	愿
猿 摇矢量预测器 .....	愿
猿 摇预测误差的量化 .....	源
猿 摇性能的比较 .....	源
猿 摇实时实现中的问题 .....	源
摇摇源 摇杂孕次厄编码方案 .....	源
猿 摇杂孕次厄系统 .....	源
猿 摇性能比较 .....	源
摇摇源 摇无记忆的蕴云量化 .....	源
猿 摇蕴云的分裂矢量量化 .....	源
猿 摇非均匀的标量量化 .....	源
摇摇参考文献 .....	源
第 缘章 摇按合成分析法(粤粤)的语音编码方案 .....	源
摇摇缘 摇按合成分析法(粤粤)的基本原理 .....	源
摇摇缘 摇通用的粤粤编码方案 .....	源
猿 摇时变滤波器 .....	源



第 1 章 语音成分的合成 .....	1
第 2 章 重建语音的产生 .....	1
第 3 章 参考文献 .....	1
第 4 章 低速率语音编码 .....	1
第 1 节 波形内插语音编码 .....	1
1.1 语音成分的波形内插 .....	1
1.2 广义窄带模型 .....	1
1.3 窄带语音编码系统 .....	1
第 2 节 正弦变换编码 .....	1
2.1 语音信号的正弦分析 .....	1
2.2 语音信号的正弦合成 .....	1
2.3 低比特率正弦编码器 .....	1
2.4 与原型波形内插相结合 .....	1
2.5 基音的频域估计法 .....	1
2.6 确定残差信号的正弦模型参数 .....	1
2.7 正弦激励参数的量化 .....	1
2.8 正弦变换解码器 .....	1
2.9 正弦变换编码器的性能 .....	1
第 3 节 参考文献 .....	1
第 5 章 宽带语音编码 .....	1
第 1 节 引言 .....	1
第 2 节 子带编码 ( 参考 [1] ) .....	1
2.1 工作原理 .....	1
2.2 子带编码的比特分配 .....	1
第 3 节 宽带语音编码 .....	1
3.1 技术要求 .....	1
3.2 方框图 .....	1
3.3 子带编码的工作原理 .....	1
3.4 子带解码器的工作原理 .....	1
3.5 正交镜像滤波器组 ( 参考 [2] ) .....	1
3.6 自适应量化与预测 .....	1
第 4 节 参考文献 .....	1
第 6 章 语音编码的方案和标准算法 .....	1
第 1 节 语音编码标准的发展概况 .....	1
第 2 节 语音编码器的属性 .....	1
2.1 比特率 .....	1
2.2 延迟 .....	1
2.3 复杂度 .....	1
2.4 语音质量 .....	1

1.1.1 规范和适应性的确认 .....	10
1.1.2 世界上有哪些标准化实体,他们是怎样开展工作的 .....	11
1.1.2.1 国际电信联盟(ITU) .....	11
1.1.2.2 北美的标准化组织 .....	12
1.1.2.3 欧洲电信标准研究所 .....	12
1.1.2.4 日本的 标准 .....	12
1.1.2.5 我国的电信标准化组织 .....	12
1.1.2.6 建立参考条款 .....	13
1.1.2.7 语音编码器的性能测量 .....	13
1.1.2.8 工作日程 .....	13
1.2 现在的语音编码标准 .....	13
1.2.1 ITU-T 公布的具有电话带宽的语音编码器 .....	14
1.2.2 ITU-T 近期发布的电话带宽语音编码器 .....	14
1.2.3 ITU-T 宽带语音编码器 .....	14
1.2.4 欧洲的数字蜂窝电话语音编码器 .....	15
1.2.5 北美的数字蜂窝语音编码器 .....	15
1.2.6 日本的数字蜂窝电话编码器 .....	15
1.2.7 国际海事卫星 语音编码器 .....	15
1.2.8 用于保密通信的语音编码标准 .....	15
1.3 语音编码器性能的总结 .....	16
1.4 附加的信息 .....	16
1.5 参考文献 .....	16
第 2 章 语音编码器的质量评价 .....	17
2.1 语音质量的定义 .....	17
2.1.1 对发音者和语言的依赖 .....	17
2.1.2 对信号功率级的依赖 .....	17
2.1.3 背景噪声 .....	18
2.1.4 多种编码的级联 .....	18
2.1.5 信道误差所产生的影响 .....	18
2.1.6 非语音信号 .....	18
2.2 语音质量的客观测量 .....	18
2.3 语音质量的主观测量 .....	19
2.3.1 绝对等级的评定试验(绝对) .....	19
2.3.2 降低等级的评定试验(比较) .....	19
2.3.3 诊断测量试验(绝对) .....	19
2.3.4 通话试验 .....	19
2.3.5 主观质量的预测 .....	19
2.4 测量语音可懂度和对发音人的识别能力 .....	19
2.4.1 可懂度 .....	19





## 第 1 章 语音编码概论

### 1.1 语音编码的数字化

语音在人类通信中具有很重要的地位，这也反映在现代技术上。电子设备是用来传送、存储、管理、识别和产生语音的工具。对于这些设备的操作，语音信号通常是表示成数字形式。这种形式对于失真来说，是相对比较顽健的，而且也便于处理。用数字表示语音信号的一个重要属性是它的比特率。它规定了一秒钟的语音需要多少比特来描述。

当传输数字语音时，所要求的带宽是比特率的函数。类似地，当存储数字信号时，比特率确定了存储媒体所需要的空间。因此，系统的价格常常是数字信号比特率的函数。语音编码的目的是要在保持重建语音质量的前提下，降低数字信号的比特率。语音编码器总是包含一个编码器和一个解码器，编码器取来原始的语音信号，并产生低速率的比特流。这个比特流传送到解码器，它能够近似地重构原始语音信号。

### 1.2 语音编码的必要性和推动力

模数转换器的输出通常是脉冲编码调制（PCM）信号。要保证电话语音的高质量（带宽限制为 300~3400 Hz），则线性 PCM 信号必须具有 8000 的采样率和 12 位量化的分辨率，所得到的比特率为 96000。这个比特率，能够作为未编码语音的参考比特率。

语音传输或存储的近似比特率，依赖于传输和存储的代价、数字语音编码（压缩）的代价和对语音质量的要求。在 1960 年以前，语音编码的高价格和低质量，使得语音编码很少被使用。由于数字信号处理硬件的效率猛增，加上语音编码研究的进展，已经明显改变了这种情况，现在语音编码已经大量使用了。

在所有的语音编码器中，重构信号和原始信号是不会完全相同的。比特率的降低是由于用较低的精度表示语音信号，或者是用语音模型的参量来表示语音信号，也能从语音信号中，除去固有的冗余度而降低比特率（无损编码）。但是，如果使用模型和参量，降低了精度就不会没有损失（有损失编码）[1]，使用降低精度的方法，来表示一个标量或矢量的过程叫做量化。在重构语音信号时，由于量化而造成的失真称为量化噪声。语音模型并不能精确地重构语音信号，甚至于模型参量不量化也会是这样。这种编码器会由于模型失真而受

到损害。在语音的编码文献中，由模型引起的失真通常也包括在“量化噪声”中，这就是广泛意义上的量化噪声。在语音编码中，对于失真的定义有很重要的意义。在某些应用中主要目标是使重构语音自然，而另外的一些应用是要求原始语音和重构语音在理解上有最大的相似性。为什么说语音编码特别困难，并具有一定的挑战性，就是很难用数学方法表示基于理解为目标的主观评价形式。

理解在语音编码中扮演了一个很重要的角色。这能够用一个简单的例子来解释。在线性传输系统中，假设动态范围没有超过，量化误差应该和语音信号的幅度无关。但是，对于小幅度信号比大幅度信号，会感觉出更多的量化噪声。利用听觉的掩蔽效应可以减少噪声，最简单的方法是用对数标度量化信号样值，即量化级之间的步长是随幅度的增加而加大。

### 语音数字编码和压缩的简要回顾

比特对数量化器（相应于  $\log_2$  的比特率），在欧洲、北美、中国和日本的电话网络中是通用的 [1]。这种技术是在 1966 年，首先由 悦悦载 (国际电报电话咨询委员会) 标准化为 悦悦载。几年后，连续可变斜率增量 (悦悦载) 调制已经介绍到军事通信中 [2, 猿]，它和 悦悦载 相似，悦悦载 调制实现后很快也普及到多方面的应用中。

1970 年左右，改进的硬件技术，把 悦悦载 和 悦悦载 变换从实际的编码操作中分开了。通常，语音信号是首先转换成 悦悦载 的信号，然后，再用语音编码算法转换成较低比特率的编码。这种方法首先应用在模拟电话网络的保密通信中。语音信号首先转换成数字比特流，然后加密，并在电话网络上用调制解调器传输。为了有更广泛的网络能够使用当时的调制解调技术，一个 悦悦载 的语音编码标准（美国联邦标准 悦悦载 [3]）就是为了这个目的而形成的。由于保密通信的需要，比较低的语音质量和产品的高价格，在当时是可以被接受的。在 1978 年，悦悦载 规定了 悦悦载 的语音编码标准 悦悦载 [3]，其目标是在通用电话网络上的应用（标准修正后称为 悦悦载）。这个编码器在低价格下提供了高质量的语音。

悦悦载 和 悦悦载 标准，是工业上广泛使用标准的一个例子，它规定了通信系统两端的设备，相互之间的可操作性。如果只是为了局部的应用，则不需要这样的标准，可以使用某些专利技术。这种专利算法的早期例子是 悦悦载 的子带编码器 [3, 4]，它用在 悦悦载 专用分支交换机 (悦悦载) 的电话线路上。这种算法以低价格提供了中等质量的语音。

在 悦悦载 和 悦悦载 制定 10 年以后，价格和性能都有了巨大的改善。语音编码器很容易在单片、低价值的数字信号处理器 (悦悦载) 上实时实现。语音编码器已经成为通信系统中的一个部件。特别重要的应用是移动通信，以及在 悦悦载 上传送语音，它对语音编码提供了强大的推动力。欧洲、北美和日本都为这个目的开发了多种语音编码标准。这些编码器的速率为 猿~ 悦悦载，属于中等比特率，而且提供了可以被接受的，比较好的语音质量。本书将详细讨论这些标准和算法。

### 语音编码器的属性

从前面部分我们已经看到，语音编码器常常是按照它的应用来开发的，要懂得为什么不同的应用就有不同的语音编码器，我们先要讨论语音编码器的属性。通常，语音编码器的主要属性是：

- 比特率；
- 主观的语音质量；

- 计算复杂度和对存储器的要求；
- 延迟；
- 对于通道误码的灵敏度；
- 信号的带宽。

本书的第 5 章要详细描述这些属性，这对于设计和开发新的语音编码器是非常重要的。

## 语音编码的声学基础知识

### 响度级和响度 [1, 2]

当我们听到一种声音时，我们常常要考虑它的“响度”。例如，我们说雷鸣是“非常响”的，而一个人在远处唱歌是“不很响”的。这些“很响”、“不很响”的定性叙述，对于某些声音已具有定量的意义。要想定量地确定某一声音的响度，最简单的方式是把它和另一标准声音相比较。1000 Hz 的纯音是选定的标准。调节 1000 Hz 纯音的声压级，使它和所研究的声音听起来有同样的响度，这个声压级就被定义为该声的响度级。响度级的单位是方 (phon)。例如，当 1000 Hz 纯音的声压级为 40 dB (相对于 20  $\mu$ Pa) 和某一方形波同样响时 (不管方形波的声压级如何)，则方形波的响度级就被认为是 40 方。

因此，声音的响度级等于 1000 Hz 纯音的声压级，单位是方 (phon) (声压级的单位是 dB)。图 5-1 是响度级与声压级和频率的关系，称为等响曲线，它反映人耳对各频率的灵敏度，其数值见表 5-1。

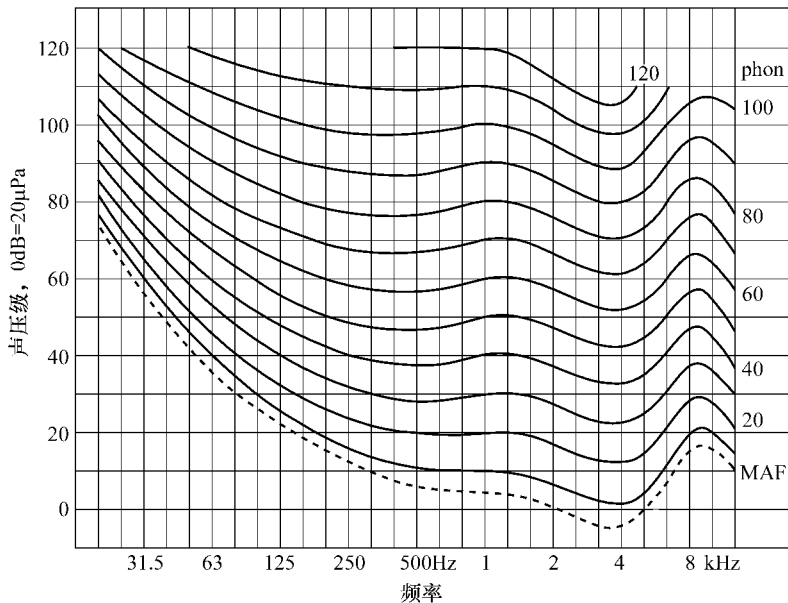


图 5-1 语音纯音的等响曲线

