



# 语音编码综述

北京邮电大学

倪维桢

**摘要:** 语音编码技术经历了半个多世纪,其编码速率已从 64kb/s 压缩到 8kb/s,并向更低速率进展。本文将回顾这一历程,并探讨今后语音编码的研究方向。

**关键词** 语音编码 线性预测 码激励

**Abstract:** Speech coding technique experiences more than half a century, the coding rate of which has compressed from 64kb/s to 8kb/s and less. This paper reviews its development course and explores future study trend.

**Keyword:** Speech coding Linear predication Code exciting

## 一、引言

自从 1937 年 A.H.Reeves 提出脉码调制(PCM)以来,开创了语音数字化通信的历程。直至今日,64kb/s 标准 PCM 系统仍占有统治地位。然而,随着微电子技术的发展,CCITT 在 80 年代初着手研究低于 64kb/s 的非 PCM 编码算法,并于 1984 年通过了 32kb/s ADPCM 的 G.721 建议。随后,于 1992 年公布 16kb

/s 低延迟码激励线性预测(LD-CELP)的 G.728 建议。最后,共轭代数码激励线性预测(CS-ACELP)的 8kb/s 语音编码 G.729 建议已在 1995 年 11 月 ITU-TSG15 全会上通过,并于 1996 年 6 月 ITU-TSG15 末次会议上通过 G.729 附件 A 减少复杂度的 8kb/s CS-ACELP 语音编解码器,正式成为国际电信标准。语音中低速率编码的

实用化取得了可喜成果。本文拟就语音编码作一简要回顾,并探讨今后语音编码的研究方向。

## 二、语音编码潜力与动力

广义来说,语音可分为三类:属于公用通信网市内和长途电话语音,可视会议电话语音和属于宽带的 CD 立体声音响。它们的带宽,标准编码和压缩编码的数值列于表 1,以便比较。

语音信号压缩编码的极限究

表 1

|           |           | 模拟带宽        | 数字速率      |         |
|-----------|-----------|-------------|-----------|---------|
|           |           |             | 标准编码      | 压缩编码    |
| 公用通信网     | 市内和长途电话语音 | 200~3 400Hz | 64kb/s    | 8kb/s   |
|           | 可视会议电话语音  | 50~7 000Hz  | 128kb/s   | 16kb/s  |
| 光盘CD立体声音响 |           | 10~20 000Hz | 1.464Mb/s | 128kb/s |

本文收稿日期:1998年5月20日

竟有多少?可以进行初步的估计。从信息论的角度来看,语音中最基本的元素可以认为是音素,英语中音素有128~256个,如果按通常说话的速率,每秒平均发出10个音素,则此时的信息率为

$$I = \log_2(256)^{10} = 80\text{b/s}$$

如果从另一角度来估值,把发音看成是以语音速率来发报文,对英语讲,每一个字母为7位码,即7bit,每分钟发125个英语单字可认为达到了通信语音速率。如果单字平均由7个字母组成,则信息率为

$$I = 7 \times 7 \times \frac{125}{60} \approx 100\text{b/s}$$

所以,可以认为语音压缩编码的极限速率为80~100b/s。当然,这时只能传送句子内容,至于讲话者的音质、音调等重要信息已全部丢失。但是,从标准编码速率到极限速率之间存在很大的跨距,这对于理论研究和实践制作无疑有着极大的吸引力。

语音压缩编码的推动力还在于:长期以来在通信网的发展中如何解决信息传输效率始终是一个关键而又极其重要的问题。有两个途径来研究这一命题。①

是研究新的调制方法与技术来提高信道的传输信息的比特率,其指标是每一赫兹带宽能传送的比特数;②是在信息源上,压缩信源的编码比特率,如标准PCM编码,对3.4kHz频带信号需要用64kb/s编码比特率传送,如压缩这一比特率显然可以提高信道传送的话路数。这对任何频率资源有限的传输环境无疑都是极为重要的,尤其是在无线通信技术决定今后通信发展命运的今天更显得重要。

实际上,语音压缩编码对话音存储、语音识别与合成等技术都直接相关。而且,语音编码技术的进展对通信新业务的发展都有着极为明显的影响,例如,话音邮件业务,实时长途翻译业务,交换机的人工智能接口技术等等。正因为此,CCITT和ITU才提出了许多紧迫需要决定的语音压缩编码标准建议以利推动通信网的发展。

### 三、语音编码方法分类

语音信号是一种时变的准周期信号。它可以近似看成由许多振幅和相位都随时间变化的正弦波构成。可以用语音的抽样波形来描述语音信号,也可以用一些语音信号的特征来描述语音信号。语音信号的特征和提取方法通常分为参数分析法和非参数分析法。采用参数法来描述语音信号,通常要首先建立语音模型,然后进行特征提取。非参数的表示则是从统计方法或变换域方法出发直接从语音信号中获得有关性质的方法。表2列出了常用的语音特征及提取方法。

利用语音信号的不同特征,编码通常可以分成两大类,分别称为波形编码和参数编码。波形编码仅仅是通过抽样和量化过程表示模拟语音波形,而参数编码则是首先把语音信号的产生用某种语音生成模型来表示,然后用语音的特征提取方法提取其必要

表 2

| 分析方法      | 参 数         | 特征描述                      |
|-----------|-------------|---------------------------|
| 短时自相关     | $\phi(m)$   | 包含谱包络以及信号细节的信息(基音)        |
| 短 时 谱     | $S(\omega)$ | 谱的细结构和包络信息                |
| 倒 谱       | $C(\tau)$   | 在倒频域将谱包络和频谱结构分开           |
| 分带谱分析     | 滤波器输出RMS值   | 能粗略描述包络信息                 |
| 过 零 率     | $Z$         | 对频谱特性粗略估值和分带谱方法结合可用于分析谱特性 |
| 分析合成法     | 模型参数        | 模型特征                      |
| 线性预测分析LPC | $\{a_i\}$   | 短时谱包络,共振频率,声道特性等          |

参数。波形编码的方法通常是利用语音信号的一些统计性质及听觉上的一些特性而对语音信号进行量化以达到压缩语音编

成大量的量化噪声从而降低了语音质量。简单的参数编码方法虽然降低语音编码速率,但合成语音质量较差。因此,介于两者

和波形编码的特点,受到广泛重视,将在第五节讨论。

前面对语音不同编码方法的特点作了简要论述,从方法上对语音编码进行了划分。还可以从比特率上对语音编码进行划分,图2表示语音编码速率的划分:高速、中速和低速。从图中可以看出编码速率划分和编码方法划分基本上是对应的。波形编码方法通常称为高速编码,其比特率在32kb/s以上,参数编码方法其速率通常在4.8kb/s以下,称为低速编码;介于中间的编码方法为中速语音编码,又称混合编码,因为它具有波形编码和参数编码的特点。

#### 四、语音编码质量评定

语音编码后,其再生语音的质量包括有清晰度(又称可懂度)与自然度。前者是衡量语音中的字、单词和句的可懂度,而后者则是对讲话人的辨识水平。这些都和人的主观听觉感知有密切关系,是一种主观判定标准,虽较为繁杂但应用较多。

##### 1. MOS分(mean opinion score)

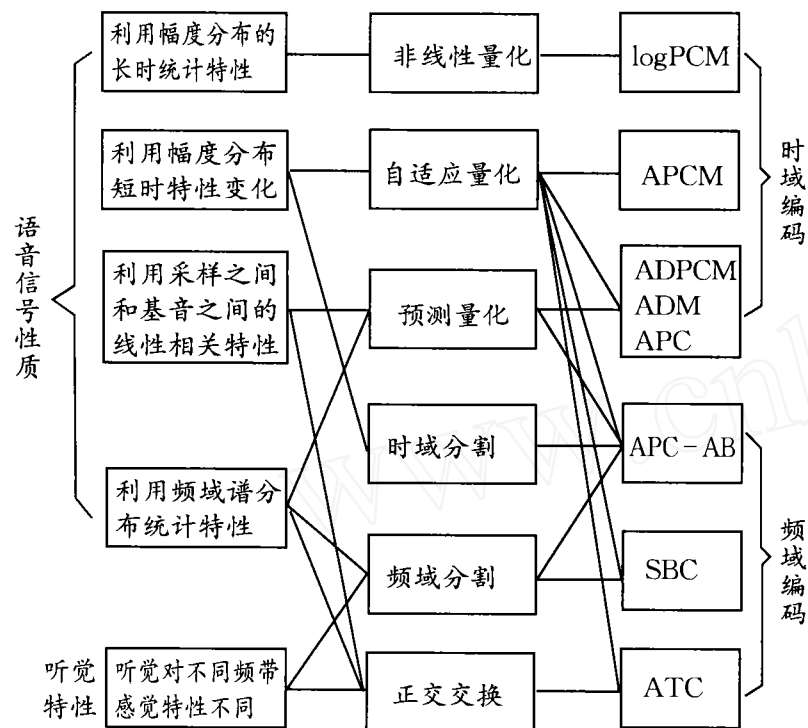


图1 不同语音波形编码方法的特点

码速率的目的。可以利用图1来表示不同的语音波形编码方法的一些特点。从图中可以看到不同编码方法利用了语音信号的不同性质。

参数编码方法则是仅仅对那些表示语音特征的参数进行编码,如LPC模型下的声道特性参数,清/浊音判决及基音信息。表3给出了波形编码和参数编码的一些特点。从表中可以看出波形编码方法不能适应于低速语音编码,这是由于粗量化造

之间的,也称为第三类编码方法——混合编码,它具有参数编码

表3

| 编码方法     | 波形编码         | 参数编码                            |
|----------|--------------|---------------------------------|
| 编码信息     | 波形           | 模型参数如LPC模型的短时谱包络,基音,清/浊音判决或其他模型 |
| 比特率      | 9.6~64kb/s   | 2.4~9.6kb/s                     |
| 语音质量评价方法 | SNR          | 谱失真和主观听音                        |
| 问题       | 随着量化粗糙语音质量下降 | 合成语音质量低,处理复杂度高                  |

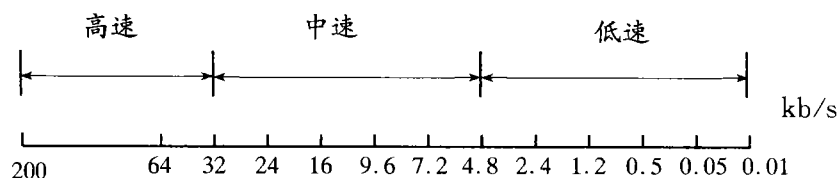


图2 语音编码比特速率划分

MOS 是一种常用的主观判定方法,以其平均主观打分来度量,用5级确定语音质量等级,分为优、良、中、差和劣5级。

MOS4.0分:为高质量语音,接近信道透明编码,常称为网络质量,再生语音质量符合长途通信要求;

MOS3.5分:为通信质量,这时语音质量会感到有所下降,但不足以妨碍正常通信;

合成语音质量,指采用参数语音综合技术再生的语音,它一般具有高的可懂度,但自然度和讲话人的确认的等级不够,这类语音一般不会超过 MOS3.0分;

高质量语音:一般频带在7kHz以上,其MOS分可达5分。

## 2. 低比特率语音的客观衡量

作为衡量语音质量的客观度量常用语音质量的补充测度,尤其是在低比特率语音编码中使用,一般有两种方法:音韵字可懂度测量 DRT(diagnostic rhyme test)和满意度测量 DAM(diagnostic acceptatilty test)。表4给出不同编码方法的以上三种测试结果,作为参考。

## 五、语音混合编码

图3给出上述3类编码方法的比特率与MOS分值的曲线。在相当一段时间内语音编码

停留在波形编码的研究上,再生语音质量没有突破性进展。随着微电子技术的迅速发展,一些复杂的算法得以硬件实现,突破了波形编码与参数编码的界线,提

出了混合编码。

得到最广泛研究的语音混合编码算法是基于线性预测技术的分析——合成编码方法,简称LPAS。一个LPAS编码器具有3个基本要素:

①基本解码器结构。解码器根据从信道发送数据确定激励信号和合成滤波器,重建语音则由激励信号作用于合成滤波器而产生。

②合成滤波器。基于LP技

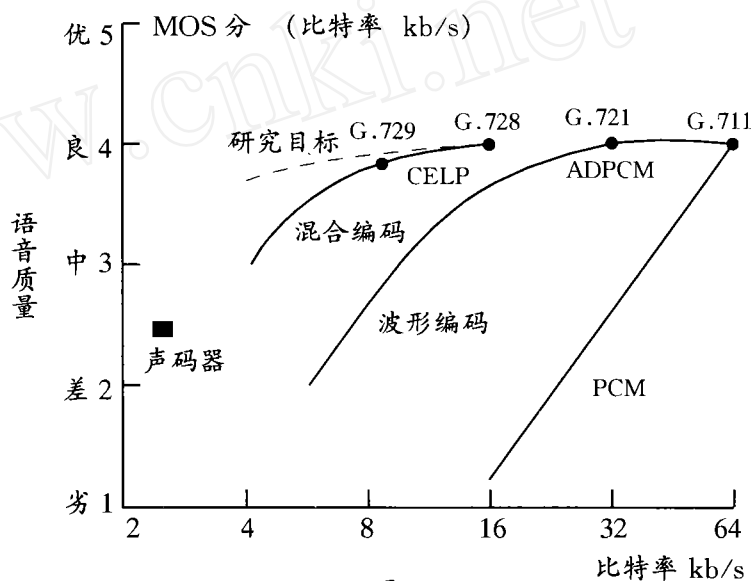


图3

表4

| 编 码 方 法                 | 测 试 方 法 |     |       |
|-------------------------|---------|-----|-------|
|                         | DRT     | DAM | MOS   |
| 64kb 8 PCM              | 95      | 73  | 4.3   |
| 32kb 8 ADPCM            | 94      | 68  | 4.1   |
| 16kb 8 LD-CELP          | 94(估值)  | 70  | 4.0   |
| 8kb 8 CELP              | 93(上界)  | 68  | 3.9   |
| 4.8kb 8 CELP(NSA1076标准) | 93(上界)  | 64  | 3.2上界 |
| 2.4kb 8 LPC(NSA1075标准)  | 90      | 55  | 2.5估值 |

术的线性时变合成滤波器通过对当前帧语音作LP分析获得的参数而周期地更新。这个合成滤波器作为整形滤波器,将一个具有平坦谱的激励信号映射成相关函数及频谱包络都与原始相似的输出信号。

③分析——合成激励编码。编码器通过将候选的激励信号送入合成滤波器的一个副本产生输出,从中选择输出信号与原始语音的感知加重失真最小的候选信号作为量化激励信号。

最早的LPAS编码思路是1981年提出来的,但最早实用的LPAS方案则是由Atal和Remede提出的多脉冲线性预测编码(MP-LPC),另外较典型的方案有规则脉冲激励线性预测编码(RPE-LPC)等。但最重要的一种LPAS算法是由Atal和Schroeder提出的码激励线性预测编码(CELP),也称随机编码,矢量激励编码(VXC)或称随机激励线性预测编码(SELP)。CELP对于MP-LPC的主要改进是采用VQ技术对激励信号编码,即激励信号以码矢量的形式存储于码本中,编码器只将最佳码矢量的地址送入信道。然而,今天一般把以LPAS为基础的用VQ技术对激励信号进行量化编码的算法统称为CELP,这时CELP已不单指一项特定的编码技术,而是一类重要的编码了。

由此可见,CELP具有波形编码和参数编码两方面特点,所以称为混合编码,它在4~

16kb/s可以得到比其他算法更高的重建语音质量,这一事实已得到公认。而且以CELP为基础的多种算法已成为国际标准,其中包括G.728建议的LD-CELP算法和G.729建议的CS-ACELP算法。

## 六、语音编码标准

1. CCITT 1972年确定,64kb/s PCM语音编码G.711建议,它已极为广泛地应用于数字通信,数字交换机以及一切语音数字化接口。

2. CCITT 1982年确定,32kb/s ADPCM语音编码G.721建议,它不仅与PCM有相同再生语音质量,而且具有比PCM更优良的抗误码性能,已逐渐广泛应用于卫星、海缆及数字语音插空设备以及可变速率编码器中,因为这种ADPCM算法在16kb/s到64kb/s的速率范围内其再生语音质量平坦,很适宜于可变速率的应用,如图3的ADPCM曲线所示。

3. CCITT 1995年5月采纳了由AT&T公司提出的一种称为低延迟码激励线性预测(LD-CELP)语音编码方案,作为16kb/s速率的国际标准,并定名为G.728建议。

LD-CELP算法特点是将语音短时谱与长时谱预测、增益因子预测等参数不是从输入语音中直接提取,而是用一50阶长的预测器在后向来实现,传送的信息则只是激励矢量,这就压缩了传输比特率。它是用分析综合方

法优化出来的5个抽样为一组的激励矢量,因此5个采样就是基本延迟,即 $5 \times 125\mu\text{s} = 0.625\text{ms}$ 。而原CCITT提出,要求编码延迟 $\leq 5\text{ms}$ 。所以,该算法远优于要求的指标,达到进入公众通信网的标准。

LD-CELP的应用领域,CCITT建议如下:①可视电话伴音;②无绳电话机;③单路单载波卫星和海事卫星通信;④数字插空设备;⑤存储和转发系统;⑥话音信息录音;⑦数字移动无线系统;⑧分组化话音。

4. 泛欧数字移动通信标准——RPE/LTP(13kb/s)

这一标准从1982年到1988年经过优选及严格试验,是从6种提案中优选并制定的。PRE/LTP是以等间隔(每3个抽样)的规则脉冲序列来模拟音源,以线性预测残差信号的振幅值为基准决定脉冲位置,残差信号,以包含长周期预测(LTP)的反馈型的量化结果为脉冲振幅。算法亦属于分析综合法。由于用帧长为20ms间隔处理,在正向提取谱包络参数与增益因子等参数,其编码延迟达25ms,对编译码处理则总延迟将达60ms。RPE/LTP编码比特率为13kb/s,若加入3kb/s纠错码保护其中重要参数,这种算法能在误码率为 $10^{-1}$ 时有足够的可懂度,比LD-CELP在同样比特速率下有更大的鲁棒性(Robust to error)。

5. 8kb/s语音编码标准

1995年11月ITU-TSG15

全会通过“共轭结构代数码激励线性预测的 8kb/s 语音编码(简称 CS-ACELP)建议 G.729。1996 年 6 月还通过了 G.729A 建议作为 G.729 的附件。

CS-ACELP 编码基于码激励线性预测(CELP)编码模型。帧长为 10ms(80 个样点)。通过对语言信号的分析,提取 CELP 模型的参数(LPC 参数、自适应和固定码本指标和增益因子)。所有这些参数被编码后传送。在解码端,这些参数用于恢复激励信号来重建语音信号。短时综合滤波器是一只 10 阶线性预测滤波器。长时或基音综合滤波采用称之为自适应码本逼近方法来实现。最后,用一个后滤波器增强重建语音质量。再生语音质量 MOS 为 3.9 分。

8kb/s 语音编码主要应用于①个人移动通信;②低 C/N 数字卫星系统:包括陆地移动卫星通信、海事卫星通信、一般航空卫星通信、稀路由和 SCPC 卫星通信系统;③高质量移动无线通信、DCME、存储/检索、分组语音和数字租用信道等。

#### 6. 4.8kb/s 语音编码标准

由于 2.4kb/s 语音编码技术(声码器)还没有突破性进展,为克服 4.8kb/s 语音编码缺点,以适用下一代保密电话。1988 年 12 月美国国防部颁布的 4.8kb/s 标准是集中全美 9 个研究单位多年研究成果,经分析比较后产生的。其质量较好,抗干扰和环境噪声能力强,清晰度测试得分

优于 16kb/s CVSD 性能,达到 DRT90, MOS3.2 分,可以和 32 kb/s 连续可变斜率增量调制质量相比较。这一标准算法仍属 CELP 编码技术。

#### 七、语音编码研究方向

近 20 年来,语音编码的研究主要集中于以 CELP 为核心的 LPAS 编码方面,在一定比特速率和相当系统复杂性条件下获得高质量的重建语音输出,并有多项实用系统和技术标准面世。随着 DSP 芯片技术的飞速进步,CELP 技术还具有一定的潜力。例如将 G.729 扩展到 6.4kb/s,用于 TDMA/CDMA 移动无线系统和 DCME。

然而,当研究活动的前沿转向 2.4kb/s 速率以下,CELP 算法即使应用更高效的量化技术也无法达到预期的指标。同时,象余弦声码器(包括 MBE 编码器及改进形式)技术更符合低速率编码的需要。另外,随着研究的深入,语音编码的研究也要求引入新的分析技术。如非线性预测、多精度时频分析技术(包括子波变换技术)、高阶统计分析技术等等。预计这些技术更能挖掘人耳听觉掩蔽等感知机理,更能以类似人耳的特性作语音的分析与合成,使语音编码系统更接近于人类听觉器官的处理方式工作,从而在低速率语音编码的研究上取得突破。当然,这一切还决定于人的听觉机理的生理学基础研究以及自动的客观测试技术发展上的突破。

#### 八、结束语

本文讨论了语音编码的方方面面,着重阐明语音编码的分类,质量评定,压缩编码的基本方法,介绍了有关语音编码的标准,对混合编码给予足够的重视,最后指出了今后语音编码的研究方向。

语音编码是一个人们熟知的领域,又是一个新颖的领域,随着信息时代的到来对语音编码提出了新的要求,语音编码技术定将进一步发展。

#### 参考文献

- 1 CCITT Draft Recommendation G.728 Coding of speech at 16kb/s using low-delay code excited linear prediction (LD-CELP), 1992
- 2 A.S.Spanias. Speech Coding: A tutorial Review, Proc. IEEE, 1994, 82(10):1541~1582
- 3 B.S. Atal and M.R. Schroeder Code-excited linear prediction (CELP): High quality speech at very low bit rates. Proc. ICASSP, 1985. 937~934
- 4 B.S. Atal High quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders. Proc. ICASSP, Tokyo, Japan, 1986. 1681~1684
- 5 J.L. Flanagan Speech Coding IEEE Trans. Commun. 1979, com-27(4):710~731
- 6 张煦. 音频处理技术进展. 数字通信, 1995, 2
- 7 李晓明. 使用共轭结构代数码激励线性预测的 8kbps 语音编码(G.729). 通讯产品世界, 1996, 11