

语音编码综述

A Review of Speech Coding

倪维桢

Ni Weizhen

提 要: 语音编码经历了半个多世纪,其编码速率已从 64kbit/s 压缩到 8kbit/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 has been studied for half a century. Its bit rates have been reduced from 64kbit/s to 8kbit/s and will go down further. This paper gives a summary of how speech coding has evolved and points out its future developments.

Key Words: speech coding, linear prediction, code excitation

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,64kbit/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 64kbit/s 的非 PCM 编码算法,并于 1984 年通过了 32kbit/s ADPCM 的 G 721 建议,随后,于 1992 年公布 16kbit/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 8kbit/s 语音编码 G 729 建议已在 1995 年 11 月 ITU-TSG15 全会通过,并于 1996 年 6 月通过 G 729 附件,正式成为国际电信标准。语音中低速编码的实用化也取得了可喜成果。本文拟就语音编码作一简要回顾。

一、语音编码的潜力与动力

语音信号压缩编码的潜力究竟有多大,其极限码速为多少?从信息论角度来估计,语音中最基本的元素可以认为是音素,语音的音素约 128~256 个,如果按通常的说话速度,每秒平均发出 10 个音素,则此时的信息率为

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

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,对英语来讲,每一个字母为七位,即 7bit,每分钟 125 个英语单字可以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

$$I = 7 \times 7 \times \frac{125}{60} = 100(\text{bit/s})$$

所以,可以认为语音压缩编码的极限速率为 80~100bit/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(64kbit/s)到极限速率(80~100bit/s)之间存在着很大的跨距(约 640 倍),这对于理论研

究和实践制作有着极大的吸引力。

推动语音压缩编码理论和技术的发展还在于通信网本身发展的需要。长期以来,在通信网发展中如何解决信息传输效率始终是一个关键而又极其迫切的问题。有两个途径来解决这一课题,其一是研究新的调制方法与技术来提高信道的传输信息比特率;其二是在信息源上压缩信源的编码比特率。这对任何频率资源有限的传输环境无疑都是极为重要的。

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。为此,CCITT 和 ITU 才提出了许多迫切需要决定的语音压缩编码标准建议,以推动通信网的发展。

二、语音编码方法分类

语音信号是一种时变的准周期信号。它可以近似地看成许多振幅和相位上都随时间变化的正弦波构成。可以用语音的抽样波来描述语音信号,也可以用一些语音信号的特征来描述语音的信号。语音信号的特征表示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,通常是首先建立语音模型,然后进行特征提取。非参数分析法的表示则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 1 列出了常用的语音特征及提取方法。

* 倪维桢 北京邮电大学教授,北京电信通信学会副理事长,北京高校成人教育研究会荣誉理事长,中国通信学会会士,原北京邮电大学副校长。多年从事数字通信、数据通信教学,进行数字信号处理、语音低速编码的研究,主编教材《数字电话原理》《数据通信原理》以及邮电百科全书(电信卷),论文 60 余篇。

表 1 常用的语音特征及提取方法

分类	分析方法	参数	特征描述
非参数法	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
参数法	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, 共振率, 声道特性

利用语音信号的不同特征, 编码通常可分成两大类, 即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码速率的目的。可以用图 1 表示不同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语音信号的不同性质。

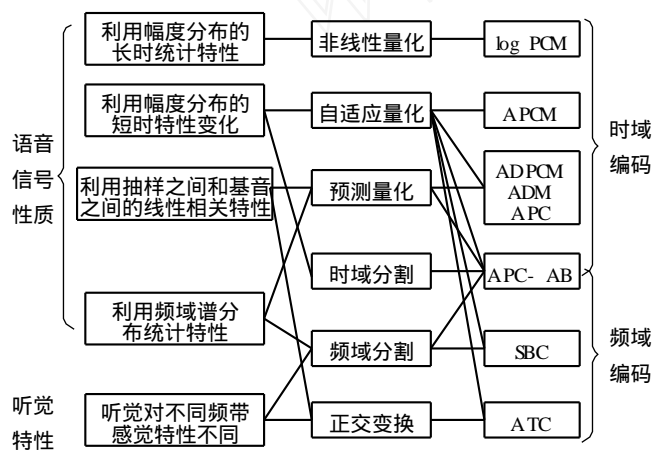


图1 不同语音编码方法的特点

参数编码方法则是仅仅对那些表征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 2 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, 但合成质量较差。因此, 介于两者之间的, 也可称为第三类编码方法——混合编码, 它具有参数编码和波形编码的特点, 受到广泛重视, 将在本文第四节讨论。

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 表示语音编码按比特速率的划分: 高速、中速和低速。从图中可以看出, 编码比特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 32kbit/s 以上, 参数编码方法, 其速率通常在 4.8kbit/s 以下, 称为低速语音编码; 介于中间的编码方法属于中速语音编码, 又称混合编码方法。

表 2 波形编码和参数编码的特点

	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	9.6~64kbit/s	2.4~9.6kbit/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1 不同语音波形编码方法的特点

图2 语音编码比特速率划分

三、语音编码质量评定

语音编码后, 其重建(再生)语音的质量包括有清晰度(又称可懂度)和自然度。前者是衡量语音中的字、单词和句的可懂程度, 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)

MOS 是一种常用的主观判定方法, 以其平均主观打分来度量, 用五级确定语音的质量等级, 分为优、良、中、差和劣五级。

MOS 4.0 分: 为高质量语音, 接近信道透明编码, 常称为网络质量, 重建语音质量符合长途通信要求; MOS 3.5 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重建的语音质量, 它一般具有高的可懂度, 但自然度和讲话人的确认等级不够, 这类语音质量一般不会超过 MOS 3.0 分。

高质量语音: 一般频带在 7kHz 以上, 其 MOS 分可达 5 分。

表 3 不同编码方法的三种测试结果

编码方法	比特速率	MOS	DRT	DAM
PCM	64kbit/s	4.3	95	73
ADPCM	32kbit/s	4.1	94	68
LD-CELP	16kbit/s	4.0	94(估值)	70
RPE-LTP	13kbit/s	3.47	—	—
SKYphone MPLPC	9.6kbit/s	3.4	—	—
CELP	8kbit/s	3.9	93(上界)	68
CELP(N.SA 1076 标准)	4.8kbit/s	3.2(上界)	93(上界)	64
STC-1	4.8kbit/s	3.5	92.7	63
LPC(N.SA 1075 标准)	2.4kbit/s	2.5(估值)	90	55

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 acceptability test)。表 3 给出了不同编码方法的以上三种测试结果,作为参考。

四、语音混合编码

图 3 给出上述三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,重建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。

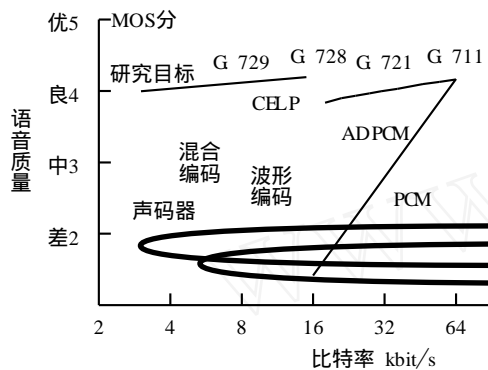


图3 三类编码方法的比特率与MOS分值的曲线

得到最广泛研究的语音编码算法是基于线性预测技术的分析——合成编码方法,简称LPA S。一个LPA S编码器具有三个基本要素。

(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语音相似的输出信号。

(3) 分析—合成激励编码器

编码器通过将候选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LPA S编码思路是1981年提出来的,但最早实用的LPA S方案则是由A tal和Remede提出的多脉冲线性预测编码(MPLPC),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LPA S算法是由A tal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)或随机激励线性预测编码(SELPC)。CELP对于MPLPC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LPA S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已不再单指一项特定的编码技术,而是一类重要的编码技术了。

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4~16kbit/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和G 729建议的CS-ACELP算法。

五、语音编码标准

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

(2) CCITT 1984确定32kbit/s ADPCM语音编码G 721建议,它不仅与PCM有相同重建语音质量,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16kbit/s到64kbit/s的速率范围内其重建语音质量平坦,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16kbit/s速率的国际标准,并定为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个抽样就是基本延迟,即 $5 \times 125\mu s = 0.625ms$ 。原CCITT要求编码延迟5ms,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:

- 可视电话的伴音;
- 无绳电话机;
- 单路单载波卫星和海事卫星通信;
- 数字插空设备;
- 存储和转发系统;
- 语音信息录音;
- 数字移动无线系统;
- 分组化话音。

(4) 泛欧数字移动通信标准——RPE/LTP(13kbit/s)。这一标准从1982年到1988年经过优选及严格试验,是从六种提案中优选并制定的。RPE/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)的反馈型量化结果为脉冲振幅。算法亦属于分析综合法。

(下转 51 页)

群至三次群跳群复接系统为例,得到的复接系统性能如下。系统的静态抖动为0.12 μ s。表3为该系统的抖动容限特性。由于测量仪器最大只能加入10 μ s的输入抖动,所以当抖动容限大于10 μ s时,无法精确测定,故以>10表示。从实验数据可知,该系统不仅满足ITU-T的有关建议,而且其指标明显优于以往的复接系统。表4为该系统的抖动转移特性,由表可知,该指标符合ITU-T的有关建议,而且系统的输出抖动几乎与输入抖动无关,使该系统具有很小的抖动积累。

表3 一次群至三次群跳群复接系统抖动容限

输入抖动 频率(Hz)	20	50	100	200	400	1k	2.4k	5k	10k	18k	40k	100k	200k												
抖动容限 (μs)	>10	7	10	5	22	3	97	3	32	3	37	3	46	3	25	3	46	3	54	3	87	2	7	1	2

表4 一次群至三次群跳群复接系统抖动转移特性

输入抖动 频率(Hz)	20	60	100	150	200	300	400	500
输出抖动(μ s)	0.122	0.121	0.122	0.122	0.121	0.122	0.122	0.122
抖动衰减(dB)	-27.8	-27.9	-27.8	-27.8	-27.9	-27.8	-28.8	-27.8

五、结束语

目前, MX2W 68231 已经过各种测试, 性能指标均符合 ITU-T 有关建议与预计结果, 并已开始应用于生产之中, 从而表明了设计方法、原理与工艺加工的正确。该芯片极大地方便了数字复接系统的小型化, 并改善了系统的性能, 增加了系统的稳定性和抗干扰性能。同时, 芯片几乎将整个复接系统集成, 充分体现了芯片系统(system on chip)的新趋势。(收稿日期: 1997-12-01)

(上接5页)

由于帧长为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13kbit/s,若加入3kbit/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在同样速率下有更大的鲁棒性(Robust to error)。

(5) 8kbit/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8kbit/s语音编码”(简称CS-ACELP)建议G.729。1996年6月还通过了G.729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)。通过对语音信号的分析,提取CELP模型的参数(LPC系数,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实现,最后用一个后滤波器增强重建语音质量。重建语音质量为MOS 3.9分。

8kbit/s语音编码主要应用于个人移动通信;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;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。

六、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LPA S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,并有多种实用系统和技术标准面世。随着DSP芯片技术的飞速发展,CELP技术还具有一定的潜力。例如将G.729扩展到6.4kbit/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2.4kbit/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的听觉掩蔽等感知机理,更能从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。

七、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。

语音编码既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 16kbit/s using low-delay code excited linear prediction (LD-CELP), 1992
- 2 Spanias A. S. Speech Coding: A tutorial Review, Proc. IEEE, 1994, 82: 1541~1582
- 3 Atal B. S., Schroeder M. R. Code-excited linear prediction (CELP): High quality speech at very low bit rates Proc. ICASSP, 1985: 973~940
- 4 Atal B. S. High quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders, Proc. ICASSP Tokyo, Japan, 1986: 1681~1684
- 5 李晓明 使用共轭结构代数码激励线性预测的8kbit/s语音编码(G.729). 通讯产品世界, 1996, 11 (收稿日期: 1997-11-19)