

文章编号:1007-9513 (2000)02-0001-04

语音编码综述

倪维桢

(北京邮电大学 电信工程学院, 北京 100876)

摘要: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,并探讨了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

中图分类号:TN912.3 **文献标识码:**A

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,CCITT在80年代初就着手研究语音压缩编码算法,并通过了一系列建议,已有将编码速率压缩到8kbit/s的国际标准(G.729建议),国家和地区级的标准则允许使用更低的码速,研究的重点正转向4kbit/s以下的低码率编码。

1 语音编码潜力与动力

语音编码的压缩极限究竟有多少?从信息论的角度看,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,对英语讲,每一个字母为7位码,即7bit,每分钟发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

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

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,从标准编码速率到极限速率之间存在很大跨距,这对于理论研究和实践制作都有着极大的吸引力。

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: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;二是在信源上压缩信源的编码率。这对于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,等等。正因为如此,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。

2 语音编码方法分类^[1,2]

语音信号是一种时变的准周期信号。语音信号可以用语音的抽样波形来描述,也可以用一些语音信号的特征来描述。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。前者,通常要首先

收稿日期:1999-12-10

作者简介:倪维桢(1933-),男,浙江海宁人,北京邮电大学教授,中国通信学会会员,多年从事数字通信、数字信号处理、语音低速编码的研究。

建立语音模型,然后进行特征提取;后者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1列出了常用的语音特征及提取方法。

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
	过零率	Z	对频谱特性粗略估值,和分带谱方法结合用于分析谱特性
参数法	分析合成法	模型参数	模型特征
	线性预测	$ a_i $	短时谱包络、共振频率、声道特性等
	分析 LPC		

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。图1表示不同波形编码方法的一些特点。参数编码方法则是仅仅对那些表征语音特征的特征参数进行编码,如 LPC 模型下的声道特性参数,清/浊音判决及基音信息。

表2给出波形编码和参数编码的一些特点。从表2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。而简单的参数编码方法虽然降低了语音编码速率,但合成的语音质量较差。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。语音编码还可从比特率上来分类,分为高速、中速和低速编码。32 kbit/s 以上称为高速编码,低于 4.8 kbit/s 称为低速编码,而 4.8 ~ 32 kbit/s 则称为中速编码。波形编码通常称为高速编码,而参数编码则为低速编码,而混合编码介于两者之间,称为中速编码。

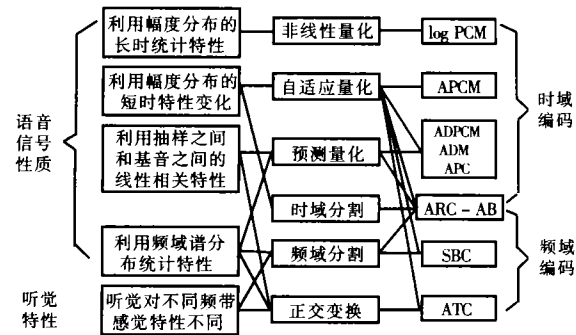


图1 不同语音波形编码方法的特点

表2 波形编码和参数编码的特点

编码方法	编码信息	比特率
波形编码	波形	9.6 ~ 64 kbit/s
参数编码	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	2.4 ~ 9.6 kbit/s
编码方法	语音质量评价方法	问 题
波形编码	SNR	随着量化粗糙语音质量下降
参数编码	谱失真和主观听音	合成语音质量低处理复杂度高

3 语音编码质量评定

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

3.1 MOS(Mean Opinion Score) 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,称为网络质量,再生语音质量符合长途通信要求;MOS 3.5 分为通信质量,这时会感到语音质量有所下降,但不足以妨碍正常通信;合成语音是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等级不够,这类语音质量,一般不超过 MOS 3.0 分;高质量语音一般频带在 7 kHz 以上,其 MOS 分可达 5 分。

3.2 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 Acceptability Test),表3给出不同编码方法的以上三种测试结果。

表3 不同编码方法的三种测试结果

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

4 语音混合编码^[3,4]

图2给出上述三类编码方法的比特率与MOS分值曲线。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,实现了混合编码。

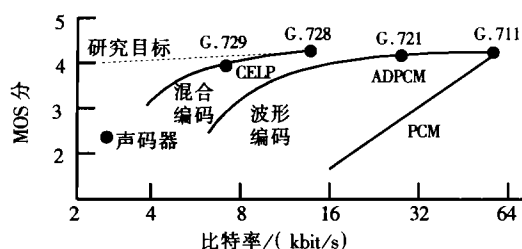


图2 三类编码方法的比特率与MOS分值的曲线

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LPAS。一个LPAS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的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。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),此外,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(Vector Quantization)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的特点,所以称为混合编码,它在4~16 kbit/s可以得到比其它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。

5 语音编码标准

5.1 64 kbit/s PCM 语音编码

CCITT 1972年确定64 kbit/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。

5.2 32 kbit/s ADPCM 语音编码

CCITT 1984年确定了32 kbit/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 16 kbit/s 到 64 kbit/s 的速率范围内其再生语音质量平坦(如图 2 的 ADPCM 曲线所示),很适宜于可变速率的应用。

5.3 低延迟码激励线性预测语音编码^[5]

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 16 kbit/s 速率的国际标准,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 个抽样就是基本延迟,即 $5 \times 125 \mu\text{s} = 625 \mu\text{s}$,原 CCITT 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 可应用于可视电话伴音、存储和转发系统、数字移动无线系统、数字插空设备、语音信息录音、分组化语音等领域。

5.4 泛欧数字移动通信标准——RPE/LTP

这一标准从 1982 年到 1988 年经过优先及严格试验,是从 6 种提案中优选制定的。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)的反馈型的量化结果为脉冲振幅。由于用帧长为 20 ms 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 25 ms,对编译码处理则总延迟将达 60 ms。RPE/LTP 编码比特率为 13 kbit/s,若加入 3 kbit/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 在同样速率下有着更大的鲁棒性(Robust to error)。

5.5 8 kbit/s 语音编码标准

1995 年 11 月 ITU-T SG15 全会通过了共轭结构代数码激励线性预测(CS-ACELP)的 8 kbit/s 语音编码的建议 G.729。1996 年 6 月还通过 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 10 ms(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来增强重建语音质量,其再生语音质量达 MOS 3.9 分。

8 kbit/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,高质量移动无线通信,DCME、存储/检索,分组语音等。

5.6 4.8 kbit/s 语音编码标准

由于 2.4 kbit/s 语音编码技术(声码器)还没有取得突破性进展。为克服 4.8 kbit/s 语音编码缺点以适用于下一代保密电话,1988 年 12 月美国防部颁布了 4.8 kbit/s 标准,它是集中全美九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,抗干扰和抗环境噪声能力强,清晰度测试得分优于 16 kbit/s 连续可变斜率增量调制(CVSD),达到 DRT 90, MOS 3.2 分,并可以和 32 kbit/s CVSD 质量相比。这一标准算法仍属 CELP 编码技术。

6 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,在一定速率和相当的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 芯片技术的飞速进步,CELP 技术还具有一定的潜力。然而,当研究活动的前沿转向 2.4 kbit/s 速率以下时,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,更能以类似人耳的特性作语音的分析与合成,使语音编码系统以更接近人类听觉器官的处理方式工作,从而在低速率语音编码(下转第 48 页)

中,不帶入实际运营局。另外,新版本、新补丁在投入运行初期不宜铺开,应选试验局进行验证,集中力量发现并解决试验局出现的问题,稳定后再推广应用。

2.5 加强信令观察、分析

实例表明,维护人员应重视用户申告。一些用户申告看似不可能,如部分局拨打苏州某用户不通,但通过对信令的分析,仍能找出一定的规律,所以维护人员一定要加强责任心,细心地分析问题。目前,信令配合问题可能出在全程全网的任何一个节点或多个节点间的配合上,信令配合障碍已不局限于某一个交换班组之内,可能牵涉到 LS、TS、转接 TS,对端 TS、LS,以及相应的信令机房,为了找到问题根源、查清障碍原因,有时需要多方配合。另外,信令

配合问题中还有一部分是活动性障碍,在分析中很难重复故障现象,因此,维护人员必须加强测试和分析,尽快找出故障原因并加以解决,确保用户通信。

3 结束语

信令配合良好是通信网正常运行的重要保证。随着通信技术的发展,新业务的不断推出,通信网将变得愈来愈复杂,不同设备、不同信令方式之间出现配合问题不可避免。但配合问题如不能得到及时解决,将影响客户的利益,影响电信网的整体运行效能,影响中国电信的形象。为此,笔者呼吁相关方面能切实重视并及时发现和解决局间信令配合问题,使电话通信网的安全可靠运行得到保证。◆

(上接第 4 页)

的研究上取得突破。当然,这一切还取决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。

参考文献:

- [1] SPANIAS A S. Speech coding: A tutorial review[J]. Proc IEEE, 1994, 82(10): 1541 - 1582.
- [2] FLANAGAN J L. Speech coding[J]. IEEE Trans Commun, 1979, 27

(4): 710 - 731.

- [3] ATAL B S, SCHROEDER M R. Code-Excited linear prediction (CELP): High quality speech at very low bit rates[A]. Proc ICASSP [C]. 1985, 937 - 940.
- [4] ATAL B S. High quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders[A]. proc ICASSP[C]. 1986, 1681 - 1684.
- [5] CCITT Draft Recommendation G.728. Coding of speech at 16 kbit/s using low-delay code excited linear prediction(LD-CELP)[S].

A Summary of Speech Coding

NI Wei-zhen

(School of Telecommunication Engineering, Beijing University of Posts and Telecommunications, Beijing 100088, China)

Abstract: This paper describes the basic method of speech coding classes, quality assessment and compressed coding, introduces the standard of speech coding, and discusses the future study of speech coding.

Key words: Speech coding; Linear prediction; Code exciting ◆