

语音编码算法综合性能比较

Integrated Performance Comparison of Speech Coding Algorithms

肖国强 张为群

(西南师范大学计算机与信息科学学院 重庆400715)

Abstract Today, speech coding is applied in many different transmission and storage applications, such as wire-line telephony, mobile and cordless telephony, internet telephony, telephone answering machines and voice output systems in user interfaces. Most of these applications ask for different design parameters, such as the bit rate, resulting in the usage of different speech coding systems. The method of assessing speech coding algorithms using analytical hierarchy process is presented in this paper.

Keywords Speech coding, AHP, Performance analysis

一、引言

语音信号处理一直在寻求一种在保持语音质量不显著下降的情况下,使语音信号的编码比特率最小的方法。利用语音信号的冗余度及某些对人耳不敏感的编码失真的知识,可以在低比特率上获得较高质量的重建语音。尽管光纤通信的出现,传输信道带宽已不是主要矛盾,但压缩编码仍一直是通信中的关键技术。这是因为对有限带宽的传输媒体(移动和卫星通信)和有限容量的存储媒体(CD-ROM 和固态存储器)的需求日益增加,因此,降低语音编码比特率就是为了低成本实现编码算法,节省有线和无线网络中的带宽,节省语音存储系统的磁盘空间。

随着超大规模集成电路(VLSI)技术的发展,特别是高速数字信号处理器(DSP)的出现,使语音编码的实时实现变得容易,复杂的语音编/解码算法能够在单片DSP上实现(如TMS320C6x)。因此,用Internet作为传输媒体的实时语音应用与日俱增,例如IP电话和视频会议系统等。国际电信联盟(ITU)的G723.1和G729建议就是为低比特率语音多媒体通信应用而制定的国际标准。

近年来,还有许多语音编码算法不断涌现出来,在实际应用中面临如何对这些算法进行选择的问题。目前,一般文献中对编码算法的性能都是从比特率、延时、复杂度和质量等方面分别进行分析,还未见到一种对各种编码算法的性能进行综合比较和评价的方法。本文试图利用T. L. Saaty提出的层次分析法^[1](Analytical Hierarchy Process, AHP)来对编码算法的性能进行综合评估。

二、编码算法的属性

语音编码器的性能常常用比特率、延时、复杂度和质量四个属性来进行衡量,因此,在分析语音编码器的性能时,主要应该考虑这些属性。值得注意的是,这些属性之间不是孤立的,而是相互紧密联系的,例如,低比特率的编码器一般比高比特率的编码器有更大的延时、更高的算法复杂度和较低的语音质量,因此在对各种编码算法进行取舍时,应根据实际应用环境,在这些属性之间进行权衡。

1. 比特率 定义为编/解码器每秒传输的编/解码语音信号的比特数。语音信号通常与其它数据共用一个通信信道,例如多媒体通信中,就同时存在语音、视频和其它形式的数字,因此,语音信号的传输不能占据全部的信道带宽,语音编码的目的就是尽可能降低语音数据的码率,以便能均衡利用信道。

在多媒体通信或Internet的混合数据传输环境中,采用自适应比特率的编码器更合适。例如,用无声压缩算法,即对活动语音采用固定比特率,对无语音期间的背景噪声采用低比特率进行编码。

2. 延时 语音编码系统的延时由三部分组成:算法延时、处理延时和通信延时。绝大部分低比特率语音编码器一次处理一帧语音数据,语音参数每帧更新和传输一次,因此,在语音数据分析之前,必须缓存一帧数据,由此导致的延时称为算法延时。

处理延时是编码器分析语音和解码器重建语音的时间,它依赖于实现编/解码器硬件的速度。算法延时加上处理延时称为单程编/解码延时。通信延时是由编/解码器之间的传输信道引起,它指一帧语音数据从编码器传送到解码器的时间。上述三部分延时的总和称

为单程系统延时,系统延时的最大值在无回音的情况下允许达400ms,然而新的测试方法表明,要达到较好的通信质量,系统延时最好不超过200ms。

3. 复杂度 算法实现的复杂度用每秒百万条指令(MIPS)、随机访问存储器(RAM)和只读存储器(ROM)三个指标来衡量。在15MIPS以下的,被认为低复杂度,在30MIPS以上的,被认为高复杂度,从系统设计者的角度看,复杂度的提高将导致更高的成本和功率消耗,在便携式应用中,较大的功率意味着降低电池的待机时间或用更大容量的电池,从而提高代价或重量。

4 质量 语音质量是四个属性中最难进行定量的,对算法的语音质量的测试,目前还没有一种真正科学的客观评价方法,因为对语音质量的最后评判是靠人的听觉,目前还是常常采用主观评价的方法。被广泛应用的主观评价方法是ATR(Absolute Category Rating),它采用5分制,1-5分别对应语音质量从低到高,最后得出平均意见分(Mean Opinion Score, MOS)。MOS在4-4.5为长途质量,3.5-4为通信质量,2.5-3.5为合成质量。表1给出了几种典型的语音编码算法的性能指标。

表1 语音编码算法的性能指标

编号	算法	比特率(kbps)	MOS	算法延迟(ms)	MIPS ^[18]	参考文献
P1	PCM (G.711)	64	4.3	0.125	0.01	[1][2]
P2	ADPCM (G.721)	32	4.1	0.125	2	[1][3][4]
P3	LD-CELP (G.728)	16	3.93	2.5	16.5	[5][6]
P4	RPE-LTP (GSM)	13	3.47	20	6	[7][8]
P5	VSELP (IS-54)	8	3.45	20	13.5	[9][10]
P6	CS-ACELP (G729a)	8	3.7	15	10.5	[11][12]
P7	ACELP (G723.1)	5.3	3.65	37.5	16	[11]
P8	CELP (FS1016)	4.8	3.16	30	16	[13][14]
P9	LPC-10e (FS1015)	2.4	2.3	22.5	7	[15][16]

三、用层次分析法综合评价算法性能

层次分析法(AHP)是美国数学家 Saaty 提出的一种整理和综合人们主观判断的客观方法,也是一种定性分析和定量分析相结合的系统分析方法。在国际上已有广泛的应用,如在资源分配、管理和政治、军事的冲突分析中。AHP 是通过分析复杂系统所包含的因素及其相互关系,将问题分解为不同的要素,并将这些要素归并为不同的层次,从而形成一个多层次的层次结构模型。语音编码算法性能比较的层次结构模型如图1所示。

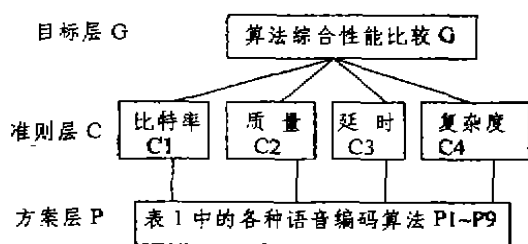


图1 层次分析结构模型

语音压缩编码最主要的目标就是在保证语音质量的前提下尽可能地降低编码速率,因此编码器的比特

率是语音编码算法的一个非常重要的属性。其次,在进行语音压缩的同时,不能使语音质量受到严重的破坏,尽量保持语音的可懂度和自然度。然而语音编码器的比特率和语音质量之间是相互矛盾的,一般高质量的语音意味着高的比特率。再次,语音编码算法的延迟在某些实际应用中对其有严格的要求,否则对语音质量将产生较大的影响,因此语音编码算法应尽量降低语音编码引起的延迟。最后,编码算法的复杂度与语音编码延迟及编码器的实现成本和功耗等方面有密切的联系,因此应采用低复杂度的算法。但是,考虑到近年来由于信号处理硬件平台取得的长足进展,单片 DSP 的运算速度和处理能力都得到大大的改善,因此,算法的复杂度相对于其它几个属性可处于比较次要的地位。然而,在便携式或移动通信设备中,由于设备的计算和处理能力受到限制,因此对算法的复杂度有一定的要求。

以下的分析是基于语音在窄带传输系统中应用的情况下,规定语音编码器的四个属性的重要性依次为比特率、语音质量、延时和复杂度。根据图1的层次分析结构模型,以表1中的各种编码算法为例,采用1-9标度法^[17]建立各层次的判断矩阵如表2(a)~(e)所示。

表2 判断矩阵及层次总排序

C1	P1	P2	P3	P4	P5	P6	P7	P8	P9	C2	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1	1/3	1/5	1/5	1/7	1/7	1/8	1/8	1/9	P1	1	1	2	4	4	3	3	5	7
P2	3	1	1/3	1/3	1/5	1/5	1/7	1/7	1/8	P2	1	1	2	4	4	3	3	5	7
P3	5	3	1	1	1/3	1/3	1/5	1/5	1/6	P3	1/2	1/2	1	3	3	2	2	4	6
P4	5	3	1	1	1/3	1/3	1/5	1/5	1/6	P4	1/4	1/4	1/3	1	1	1/2	1/2	2	4
P5	7	5	3	3	1	1	1/3	1/3	1/4	P5	1/4	1/4	1/3	1	1	1/2	1/2	2	4
P6	7	5	3	3	1	1	1/3	1/3	1/4	P6	1/3	1/3	1/2	2	2	1	1	3	5
P7	8	7	5	5	3	3	1	1	1/2	P7	1/3	1/3	1/2	2	2	1	1	3	5
P8	8	7	5	5	3	3	1	1	1/2	P8	1/5	1/5	1/4	1/2	1/2	1/3	1/3	1	3
P9	9	8	6	6	4	4	2	2	1	P9	1/7	1/7	1/6	1/4	1/4	1/5	1/5	1/3	1
$\lambda_{max} = 9.498$ RI=1.45 CR=0.043 (a)										$\lambda_{max} = 9.215$ RI=1.45 CR=0.019 (b)									
C3	P1	P2	P3	P4	P5	P6	P7	P8	P9	C4	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1	1	8	9	9	9	9	9	9	P1	1	9	9	9	9	9	9	9	9
P2	1	1	8	9	9	9	9	9	9	P2	1/9	1	7	3	6	5	7	7	3
P3	1/8	1/8	1	6	6	5	8	7	6	P3	1/9	1/7	1	1/5	1/2	1/3	1	1	1/5
P4	1/9	1/9	1/6	1	1	1/2	3	2	1	P4	1/9	1/3	5	1	4	3	5	5	1
P5	1/9	1/9	1/6	1	1	1/2	3	2	1	P5	1/9	1/6	2	1/4	1	1/2	2	2	1/3
P6	1/9	1/9	1/5	2	2	1	4	3	2	P6	1/9	1/5	3	1/3	2	1	3	3	1/3
P7	1/9	1/9	1/8	1/3	1/3	1/4	1	1/2	1/3	P7	1/9	1/7	1	1/5	1/2	1/3	1	1	1/5
P8	1/9	1/9	1/7	1/2	1/2	1/3	2	1	1/2	P8	1/9	1/7	1	1/5	1/2	1/3	1	1	1/5
P9	1/9	1/9	1/6	1	1	1/2	3	2	1	P9	1/9	1/3	5	1	3	3	5	5	1
$\lambda_{max} = 9.866$ RI=1.45 CR=0.075 (c)										$\lambda_{max} = 9.762$ RI=1.45 CR=0.066 (d)									
G	C1	C2	C3	C4	W	P	C1	C2	C3	C4	层次总排序 W								
C1	1	3	4	5	0.5423	P1	0.0165	0.2351	0.3199	0.4478	0.1464								
C2	1/3	1	2	3	0.2333	P2	0.0260	0.2351	0.3199	0.1785	0.1288								
C3	1/4	1/2	1	2	0.1397	P3	0.0481	0.1529	0.1475	0.0244	0.0844								
C4	1/5	1/3	1/2	1	0.0847	P4	0.0481	0.0606	0.0374	0.1050	0.0543								
$\lambda_{max} = 4.051$ RI=0.90 CR=0.019 (e)						P5	0.0960	0.0606	0.0374	0.0387	0.0747								
						P6	0.0960	0.0970	0.0567	0.0561	0.0874								
						P7	0.1928	0.0970	0.0186	0.0244	0.1319								
						P8	0.1928	0.0400	0.0252	0.0244	0.1195								
						P9	0.2837	0.0216	0.0374	0.1008	0.1726								
						$RI=1.45$ CR=0.030 (f)													

对各判断矩阵进行一致性检验,判断矩阵的一致性指标是用以检查人们判断思维的一致性,定义为

$$CI = (\lambda_{max} - n) / (n - 1)$$

其中, λ_{max} 和 n 分别为判断矩阵的最大特征根和阶数。由于随机干扰, 还需引入判断矩阵的平均随机一致性指标 RI, RI 的数值可查表得出^[17]。把 $CR = CI/RI$ 定义为随机一致性比率, 根据大量实践的结果, 当 $CR < 0.1$ 时, 便可认为判断矩阵具有满意的一致性。采用 AHP 计算中的和积法计算出各判断矩阵的最大特征根及其所对应的归一化特征向量, 此特征向量也就是层次单排序中各种方案相对重要性的排序权值, 同时得到各个判断矩阵一致性检验结果见表2所示, 各个层次单排序的权值及层次总排序的分析结果列于表2(f)中。

结论 从表2(f)层次总排序的分析结果可以得出几点结论: (1)各种编码算法通过相应的权值大小定量地反应出其综合性能的优劣程度, 其中 P9(LPC-10e)的权值最大, 在窄带传输系统要求极低比特率的情况下, 也就是比特率相对于其它几个属性有较大的权重时, P9是应该首先考虑的编码方法; (2)P1(PCM)和 P2(ADPCM)都属于波形编码, 虽然其传输比特率较高, 但是其它几个属性的指标都较高, 因此其综合性能的权值也较大, 与低比特率编码算法 P7(ACELP)的性能相当; (3)从 P1、P2、P3权值的比较可以看出, 比特率的降低并不意味着综合性能一定提高, 还与其它属性的指标有关; (4)对于同类型的 P3~P9参数编码算法, 除了比特率外, 其它指标相差不是很大, 因此其总体性能的权值随着比特率的降低几乎逐渐增加。

参考文献

- 1 Jayant N S. Coding Speech at Low Bit Rate. IEEE Spectrum, 1986, 8: 58~63
- 2 Jayant N S, Noll P. Digital Coding of Waveforms. Prentice-Hall, INC. Englewood Cliff, NJ 1984
- 3 Benevuto N, et al. The 32Kb/s Coding Standard. AT&T Technical Journal, 1986, 65(5): 12~22

(下接第124页)

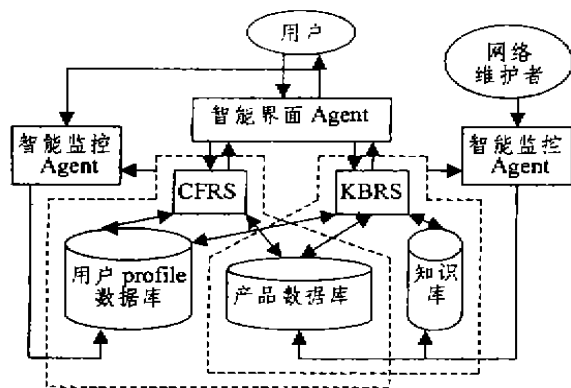


图1

图1中所示的为经过修改后的混合推荐结构。它和 Thomas Tran 等提出的混合式推荐系统结构相比较,主要区别在于:1)采用智能监控 Agent 来进行库的建立与维护。2)拓展 KBRS。在 KBRS 中加入用户模型,即用户数据库的部分内容。

结语 现有的大量商务网站对人机交互时用户信息的过于依赖其出发点是为了增强整个系统的有效性,如果通过智能 Agent 在基于 Web 的交互中能提炼出相关的用户模型和消费习惯无疑对整个系统的性能

是有益的,但对于智能 Agent 而言,相关的推理能力及学习算法的优劣又成为整个系统的关键所在。这正是后续研究的目的所在。

参考文献

- 1 Gini M. Agent and other "Intelligent Software" for e-Commerce. CSOM, University of Minnesota. <http://www-users.cs.umn.edu/~gini/csom.html>
- 2 <http://www.amazon.com/>
- 3 Burke R. Knowledge-based Recommender System. In: the Encyclopedia of Library and Information Science
- 4 <http://personallogic.com/>
- 5 Burke R. Intelligent Knowledge-based and Collaborative-filtering Recommender system. In: Proceedings of the workshop on AI and Electronic Commerce. AAAI 99, Orlando, Florida, 1999
- 6 Tran T, Cohen R. Hybrid Recommender System for Electronic Commerce. In: the AAAI-2000 Workshop on Knowledge-based Electronic Markets, July 31, Austin TX, USA
- 7 Towle B, Quinn C. Knowledge-based Recommender System Using Explicit User Models. In: the AAAI-2000 Workshop on Knowledge-based Electronic Markets, July 31, Austin TX, USA

(上接第127页)

- 4 CCITT Recommendation G. 721. 32Kb/s Adaptive Differential Pulse Code Modulation (ADPCM). 1988
- 5 CCITT Draft Recommendation G. 728. Coding of Speech at 16Kb/s Using Low-Delay Code Excited Linear Prediction (LD-CELP). 1992
- 6 Chen J. A Low-Delay CELP Coder for the CCITT 16KB/s Speech Coding Standard. IEEE Trans. on Selected Areas in Communication, Special Issue on Speech and Image Coding, Ed. N. Hubing, 1992, 6: 830~849
- 7 GSM 06.10. GSM Full-Rate Transcoding. [Technical Rep. Vers. 3.2]. ETSI/GSM, 1989
- 8 Vary P, et al. Speech Code for the European Mobile Radio System. Proc. ICASSP'88, 1988
- 9 EIA/TIA-PN2398 (IS-54). The 8 kbits/s VSELP Algorithm. 1989
- 10 Gerson I, Jasiuk M. Vector Sum Excited Linear Prediction (VSELP) at 8 kbits/s. Proc. ICASSP'90, 1990: 205~208
- 11 <http://www.robotics.net/papers/integratedvoice.html>
- 12 Salami R, et al. ITU-T G. 729 Annex A: Reduced Complexity 8 Kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data. IEEE Communication Magazine, 1997 (9): 56~63
- 13 Campbell J, et al. The Proposed Federal Standard 1016 4800bps Voice Coder: CELP. Speech Technology, 1990, 4: 58~64
- 14 Campbell J P, Welch V C, Tremain T E. An Expandable Error-Protected 4800 BPS CELP Coder (U.S. Federal Standard 4800 BPS Voice Coder). IEEE ICASSP, 1989: 735~737
- 15 Supplee L M, Cohn R P, Collura J S, McCree A V. MELP: The New Federal Standard at 2400 bps. IEEE ICASSP, 1997
- 16 Tremain T E. The Government Standard Linear Predictive Coding Algorithm: LPC-10. Speech Technology 1982, 4: 40~49
- 17 赵焕臣, 许树柏, 金生编著. 层次分析法——一种简易的新决策方法. 北京: 科学出版社, 1986.
- 18 <http://www.ti.com/sc/docs/apps/telecommunications/celp-coder.html#App-Notes>