

文章编号: 1001-0920(2003)04-0475-04

VoIP 的语音质量分析与控制

黄永峰, 李 星

(清华大学 电子工程系, 北京 100084)

摘 要: 分析了 VoIP 语音质量的影响因素, 通过 E 模型定量地描述了语音质量与端到端延迟和丢包率的关系。为了控制 VoIP 的语音质量, 计算出 VoIP 系统在各种情况下的语音质量极限, 提出一种自适应编码和分组封装的控制策略。将该方法应用于自行开发的 IP 电话网关, 实际测试证明能在很大程度上提高 VoIP 的语音质量。

关键词: VoIP; 语音质量; 编码器

中图分类号: TP393

文献标识码: A

Analysis and control of speech quality in VoIP

HUAN G Yong-feng, LI X ing

(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Factors affecting the speech quality of VoIP system are analyzed. A mathematic relation between speech quality and end-to-end delay and/or rate of packet loss is presented based on E model. To control the speech quality of VoIP, the limit of speech quality is calculated in all cases and some self-adaptive encode and package methods are suggested. These methods are applied to IP telephony gateway and the results of testing show that these adaptive mechanisms can improve the speech quality of VoIP greatly.

Key words: VoIP; Speech quality; Codec

1 引 言

语音与 IP 技术的结合, 迎来了新的网络环境, 即 VoIP。这种新环境提供了低成本、高灵活性、高效率的增强应用。IP 在硬件、软件和网络协议等方面的提高和发展, 进一步推动了新的集成化基础设施的迅速发展。然而 VoIP 与传统电话语音相比, 也有其致命的弱点, 即语音质量较差, 这已成为 VoIP 研究领域的一个难题。令人欣喜的是, 随着 IP 技术的发展, VoIP 服务质量(QoS)的研究取得了不少成果, 提出了许多改进 QoS 的策略。归纳起来可以分为三类^[1]: 注册许可控制策略、资源预留策略和自适应策略。前两类策略分别采用资源预留(RSVP)等

协议和改变 IP 网络的结构, 以改善 VoIP 的语音质量, 但这两类策略的实施涉及到整个网络环境的改造, 实现起来困难较多; 后一类策略是在 VoIP 的终端产品(如 IP 电话网关)中, 采用一些能根据网络状态变化而进行相关参数调整的技术, 因此实现起来较为简便。

本文主要介绍作者在开发 IP 电话网关时, 所采用的自适应编码和分组封装策略。文中分析了影响 VoIP 语音质量的主要因素, 并根据 E 模型讨论了 VoIP 语音质量与端到端延迟和丢包率之间的关系, 提出了编码和分组封装的自适应调整方法和质量控制策略。实验证实了本文所提出方法能在很大程度

收稿日期: 2002-03-08; 修回日期: 2002-08-12。

作者简介: 黄永峰(1967—), 男, 湖北赤壁人, 博士后, 从事计算机网络、通信技术等研究; 李星(1956—), 男, 北京人, 教授, 博士生导师, 从事计算机网络、多媒体信息处理技术等研究。

上改进语音质量。

2 VoIP 语音质量的影响因素

2.1 语音质量的评估参数

语音质量是指复制语音的保真度和可理解性,即从对话中提取信息的能力^[1]。在VoIP网络中,衡量语音质量的参数主要有以下几个:

1) 清晰度:指从会话过程中可提取出的信息量。清晰度可通过感知语音质量测量(PSQM)或平均判分(MOS)等方法来测量。

2) 端到端的延迟:指可能影响人们感受的会话延迟。在不加控制的情况下,可能导致经常性中断。

3) DTMF音调分析:如果使用低比特率语音编解码器,DTMF音调就不能正确再现。DTMF重要的失真参数包括放大和频移。

2.2 VoIP 语音质量的影响因素

影响VoIP语音质量的因素可分为两类,即语音传输因素和网络条件因素。具体表现为以下几点:

1) 压缩编码:任何压缩算法都会对语音质量有所影响。忽略所有其他衰减原因,语音质量只能达到压缩算法在接收站准确重构语音信号的水平。如果需要很高的语音质量,则意味着只能使用64 kbit/s PCM编码^[2]。

2) 语音活动检测(VAD):是通过自动检测会话中的静音时间段,并在这些时间段暂停数据流的产生,以减少发送的语音数据量。VAD的技术难点在于如何正确标识讲话何时停止,又何时重新开始。由于静音判断的失误,使用VAD技术常会产生所谓的“切尾”现象。

3) 回声:是由语音网络中的电气反射引起的。延时超过16 ms的回声对说话人会有影响。

4) 延时:是直接影响语音质量的主要因素。过长的延时会导致讲话人声音产生重叠和回声,更长的延时甚至会使交谈无法进行。

5) 延迟抖动:固定的延迟将干扰谈话和回答的节奏,变动的延迟(简称抖动)将在发音之间产生随机中断。由于排队延时、可变的分组大小、中间链路和路由器上相对负载等因素的变化,IP网络不能提供稳定的性能,常常引起分组到达延迟的变化^[3]。

6) 分组丢失:IP网络将丢失分组作为解决网络拥挤的手段,当语音经过IP网络时,产生丢包是难免的。对于数据,如要恢复丢失的分组,发送端只需简单地重新发送丢失的数据并降低它们的通信速率;对于语音,由于实时性的要求,没有时间进行重新发送丢失语音,只能由终端设备采用分组丢失恢

复算法来解决。

3 VoIP 语音质量的分析

根据ITU-T G.114和G.131建议,端到端延迟对非失真语音的影响是不明显的。但是如果语音一旦出现失真,则端到端延迟对语音质量会有显著的影响,而且延迟大小的不同,其影响也不同。特别是对于VoIP系统,由于压缩编码等因素的存在,失真不可避免。因此,提高VoIP语音质量的关键是同时控制端到端延迟和分组丢失。尽管人们对VoIP的语音呼叫可以忍受一定的失真,但这种容忍是有一定界限的。如何确定这种界限?下面采用E模型进行分析。

3.1 E模型

为了分析电话网络端到端的语音质量,ITU-T G.107建议提出一种E模型的分析工具^[2]。E模型认为,电话的语音质量是由语音的传输参数决定的,并将这些传输参数所产生的质量损耗综合在一个参数中,用 R 表示,即

$$R = R_0 - I_s - I_d - I_e + A \quad (1)$$

其中: R_0 指噪音带来的影响; I_s 包括语音量化和压缩编码所产生的质量损耗; I_d 指回声延迟所造成的质量损耗; I_e 包括由于使用特殊设备而产生的质量损耗,或由分组丢失带来的影响; A 是预期因素,表现为参数 R 的减小,即由于VoIP系统比传统PSTN电话更具“接入优越性”,使用户愿意接受质量损耗的存在。

参数 R 可用来预测主观用户反应,比如平均判分(MOS)的百分比。在上述参数中,影响 R 的关键参数是 I_d 和 I_e ,即端到端延迟和失真。ITU-T G.109建议草案认为,参数 R 的取值范围[90, 100], [80, 90], [70, 80], [60, 70], [50, 60]分别对应于呼叫质量很高、高、中、低、差5个级别。参数低于50表示该呼叫质量无法接受。对于PSTN电话,用户能接收的呼叫质量的 R 值不低于70^[3]。

由式(1)可知,即使两个呼叫的参数 R 相同,但可能给人的主观感觉完全不同。因为一个呼叫可能产生清晰的未失真的语音(如 $I_e = 0$),但会有相当大的延迟(如 $I_d = 10$);而另一个呼叫可能有轻微的语音失真(如 $I_e = 10$),但它的延迟却不会引起注意(如 $I_d = 0$)。虽然原因有所不同,但E模型会判断这两个呼叫具有相同的MOS值。

3.2 端到端延迟的影响

质量损耗值 I_d 是由回声和延迟引起的。通信双方如果第1方听到的语音信号产生延迟, R 就要减

少 I_{d_0} 。该质量损耗值等于 3 个质量损耗值之和,即

$$I_d = I_{d_1} + I_{d_2} + I_{d_3} \quad (2)$$

其中: I_{d_1} 表示说话者的回声质量损耗值,该回声是在接近第 1 方处由反射产生的,这个回声的水平在很大程度上受靠近第 2 方的回声损耗 E_{L_2} 的影响; I_{d_2} 是收听者的回声质量损耗值,收听者的回声也干扰第 1 方,第 1 方听到的是来自第 2 方的原始信号并伴随着该信号的衰减回声,这个回声是在靠近第 1 方处由反射决定的,并由回声损耗 E_{L_1} 产生衰减,然后在接近第 2 方处再次产生反射,并由 E_{L_2} 产生衰减; I_{d_3} 是延迟对交互谈话的损耗,如果端到端延迟过大,则交互对话将无法进行。

回声损耗的典型值为 21 dB,如果使用回声消除器,则可增加回声损耗值 E_{L_1} 和 E_{L_2} ,一般可增加到 30 dB,甚至更大。根据 E 模型,可计算出未失真语音条件下端到端延迟对参数 R 在不同回声损耗值的影响。对于 VoIP,未失真的呼叫是使用 G 711 格式编码并在传送中没有帧丢失的呼叫。ITU-T G 114 和 G 131 建议要求确保 PSTN 呼叫的参数 R 值不得少于 70。对于 VoIP 而言,如果 R 值也按 70 计算^[4],则对于 21 dB 的回声损耗,端到端延迟最大只能为 25 ms;如果对呼叫进行很好的回声消除,则端到端延迟最大可达 400 ms。

3.3 失真的影响

如果接收方听到的语音信号失真,则参数 R 将减小 I_e 数量。作为失真质量损耗值, I_e 的大小取决于两个因素:发送方的语音压缩编码算法和语音分组在传输中的丢失率。不同的压缩编码算法所造成的失真质量损耗是不同的,表 1 列出了 VoIP 常用的标准编解码器对应的失真质量损耗值和固有质量值。其中:编码的固有质量定义为端到端延迟为 0,丢包率为 0 时参数 R 的最大值^[5]。

表 1 标准编解码器的主要参数

标准	类型	速率	算法延迟	I_e	固有质量
G 711	PCM	64	0.125	0	94.3
G 729 A	CS-ACELP	8	15	10	84.3
G 723 1	MP-MLQ	6.3	37.5	15	79.3
	ACELP	5.3		19	75.3

4 自适应编码和分组封装策略

由对 E 模型的分析可知,VoIP 的语音质量主要与端到端延迟和失真损耗有关。在设计一个 VoIP 系统时,通过控制端到端的延迟和丢包率,可在一定程度上控制语音质量。尽管 VoIP 系统的丢包率和延迟与整个网络环境有关,但在设计 VoIP 的终端设备时(如 IP 电话网关),通过采用不同的编码算法,不同的分组封装策略和不同的分组丢失恢复机制,可以控制 VoIP 系统的丢包率和端到端延迟。

4.1 端到端延迟的控制

由 IP 电话网关组成的典型 VoIP 系统的端到端延迟如图 1 所示。根据图 1,系统的端到端延迟 $D_{end-end}$ 可表示为

$$D_{end-end} = D_{IP} + 2(D_{gateway} + D_{PSTN} + D_{tel}) \quad (3)$$

其中:电话延迟 D_{tel} 和 PSTN 延迟 D_{PSTN} 不是很显著,而网关延迟 $D_{gateway}$ 对端到端延迟影响较大,但却很容易控制。因为 $D_{gateway}$ 是对网关语音进行处理和对分组进行封装所带来的延迟,因此可表示为

$$D_{gateway} = D_{pack} + D_{playout} \quad (4)$$

其中: $D_{playout}$ 表示分组的播放延迟,取决于延迟抖动缓冲区的深度; D_{pack} 表示分组封装延迟,可表示为

$$D_{pack} = N D_{encode} \quad (5)$$

其中: N 为分组封装语音帧的数目,可取 1, 2, ...; D_{encode} 为压缩编码的算法延迟,对于 G 723 1 为 37.5 ms。由式(4)和(5)可知,通过改变分组包含语音帧的数目 N ,可在一定程度上调整 VoIP 的端到端延迟,进而控制 VoIP 的语音质量。

4.2 丢包率的控制

欲控制 VoIP 的丢包率,可采用各种语音帧插

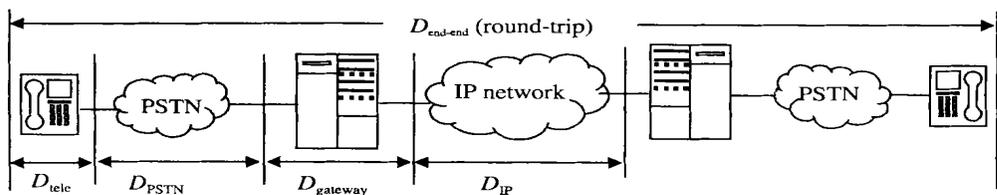


图 1 VoIP 系统的延迟组成

值和产生低速率冗余分组的方法。作者在开发 IP 电话网关时,采用表 2 所示的低速率冗余分组的封装方法,以控制 VoIP 的丢包率。

表 2 不同编码和分组封装格式

编号	标准	冗余度	分组/帧	$D_{\text{pack}}/\text{ms}$
1	G 723 1/6 3	0	1	30
2	G 723 1/5 3	0	1	30
3	G 711/64	0	—	20
4	G 723 1/6 3	1	2	30
5	G 723 1/5 3	1	2	30
6	G 729 A/8 0	1	2	10
7	G 723 1/6 3	0	2	60
8	G 723 1/5 3	0	2	60
9	G 729 A/8 0	0	2	20
10	G 711/64	0	—	60
11	G 723 1/6 3	2	4	60
12	G 723 1/5 3	2	4	60
13	G 729 A/8 0	2	4	20
14	G 729 A/8 0	0	4	40

4.3 自适应语音质量控制策略

前已引入编码算法的固有质量的概念,在此给出损耗冗余量的概念。损耗冗余量定义为编码器的固有质量与用户能接受的最低语音质量之差。只要编码器存在一定的损耗冗余量,便允许该编码器语音传输的端到端延迟超过 150 ms,或允许发生一定的分组丢失,而且分组丢失与端到端延迟可以交替出现。如果 VoIP 的最低语音质量按 70 计算,则可根据 E 模型计算出各种编码器在端到端延迟低于 150 ms 时可容忍的最大分组丢失率,计算结果如表 3 所示。同时,也可计算出每种编码器在无分组丢失时可容忍的最大端到端延迟,如表 4 所示。这两种极限值可作为设计 VoIP 系统语音质量控制的依据。

由上述分析可知,VoIP 终端中的编码器在选择编码算法和进行分组封装时,可按表 2 选择不同冗

表 3 可容忍的分组丢失的界限

标准	编码速率	分组丢失极限
G 711	64	10
G 729 A + VAD	8	3.4
G 723 1 + VAD	6.3	2.1

表 4 可容忍的端到端延迟的极限

标准	编码速率	分组丢失极限/ms
G 711	64	400
G 729 A + VAD	8	296
G 723 1 + VAD	6.3	253

余信息和帧数来控制 VoIP 语音质量。当 VoIP 系统的丢包率较小时,编码器自动选择帧数 N 较多的分组封装方式来发送语音分组,如表 2 中编号为 8,9 等分组格式,这样能在保证语音质量的前提下提高带宽的利用率。当 VoIP 网络的传输延迟较大时,则选择帧数 N 较少且带有冗余帧的分组封装方式来发送语音分组,如表 2 中编号为 4,5 等分组格式,保证语音质量不低于 70。参考表 3 和表 4,可在丢包率与延迟之间找到最佳平衡点,只要使二者所带来的质量损耗不超过该编码器的损耗冗余量即可。

4.4 实验测试

作者采用上述控制策略开发出一个能处理 E1 的 IP 电话网关,并将该网关连接在 CERNET 环境下,组成一个 VoIP 实验系统,跟踪测试系统的延迟、丢包率及语音质量等参数。编码器在采用 G 729 A 和 G 723 1 的情况下,DTMF 传输频移小于 ± 1.2 Hz,相位抖动小于 7°,MOS 值大于 3.6,延迟小于 180 ms,丢包率小于 3%。测试结果表明,本文所采用的自适应控制策略能在很大程度上保证 VoIP 的语音质量。

5 结 语

本文通过 E 模型分析了 VoIP 的语音质量与延迟和丢包率之间的关系,提出一种自适应编码和分组封装策略,并将该策略成功地应用于 IP 电话网关。实际测试证实,上述策略能在很大程度上提高 VoIP 系统的语音质量。

参考文献(References):

- [1] Chin K V, Hui S C, Foo S. Enhancing the quality of internet voice communication for internet telephony systems[J]. *J. of Network and Computer Applications*, 1998, 21: 203-218
- [2] Richard V Cox, Peter Kroon. Low bit-rate speech coders for multimedia communication[J]. *IEEE Communication Magazine*, 1996, 34(12): 21-27.
- [3] Huang C M, Lee R Y. Quantification of quality-of-presentation (QOPs) for multimedia synchronization schemes[J]. *Computer Communication*, 1997, 12(3): 76-102
- [4] Campbell A, Coulson G, Huchison D. Transporting QoS adaptive flows[A]. *Multimedia Systems[C]*. Berlin: Springer-Verlag, 1998: 167-178
- [5] Ranjee R, Jim K, Don T, et al. Adaptive playout mechanisms for packetized audio applications in wide-area networks[A]. *Proc IEEE IN FOCOM 94[C]*. Toronto, 1994: 680-688