

# 一种 VoIP 语音质量评价模型

胡治国<sup>1,2</sup> 张大陆<sup>1</sup> 张俊生<sup>1</sup>

(同济大学计算机科学与技术系 上海 201804)<sup>1</sup> (中国人民解放军 65583 部队 辽阳 111000)<sup>2</sup>

**摘要** 在 VoIP 系统中,传输网络性能(QoS)参数对可感知语音质量(Quality of Experience, QoE)起着基础性的影响作用,但 QoS 取值情况并不能直接反映和代表 QoE 水平。为此,基于对 VoIP 传输特征的分析,首先采用 PESQ, E-Model 算法分析了单个 QoS 参数对 QoE 损伤的影响;在单个因素计算的基础上,通过对 E-Model 算法的扩展研究了 QoS 参数综合作用情况下语音 QoE 值的变化情况;采用回归分析的方法建立了 QoS 参数与语音 QoE 的映射模型,模型构成简单。验证实验表明,该模型与语音 QoE 客观评价方法之间具有很高的相关度,满足对网络运行状况及 VoIP QoE 实时监测的要求。

**关键词** 服务质量,体验质量,语音质量客观评价,回归模型

**中图法分类号** TP393 **文献标识码** A

## Perceived Speech Quality Estimation Model for VoIP-based Networks

HU Zhi-guo<sup>1,2</sup> ZHANG Da-lu<sup>1</sup> ZHANG Jun-sheng<sup>1</sup>

(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)<sup>1</sup>

(The Chinese People's Liberation Army, 65583 Troops, Liaoyang 111000, China)<sup>2</sup>

**Abstract** In VoIP system, the performance of network transmission makes a fundamental impact on quality of experience. However, the traditional research based QoS can not be directly reflected to be Quality of Experience. We combined ITU PESQ and E-model algorithm and made a comprehensive analysis of the impact on user experience which was caused by packet loss, end-to-end delay, jitter, voice coding, and so on. Our contributions are thus twofold: firstly, we offered a detailed analysis of the impact of these parameters and their interactions on the perceived conversational quality. Secondly, we adopt regression model for single-ended(nonintrusive) method for objective speech quality estimation according to different conditions. The experimental results show that the objective assessments are correlated well with objective assessments ones, which indicates that the method proposed here is effective.

**Keywords** QoS, Quality of experience, Objective assessment of speech quality, Regression model

## 1 引言

按照评价主体及方法的不同,通常可感知语音质量(QoE)评价分为主观评价、客观评价、间接评价<sup>[1]</sup>。主观评价是以人为主体,根据某种预先约定的指标来对失真语音或参考原始语音划分质量等级。具有代表性的方法包括:平均意见分 MOS<sup>[2]</sup>、判断韵字测试方法 DRT<sup>[3]</sup>、失真平均意见分 DMOS<sup>[4]</sup>等。主观评价的优点在于直接、准确并易于理解。然而,主观评价不但对听评条件与流程有严格要求,还需要对大量听评者的评价结果做出统计,因此主观评价成本高、灵活性差<sup>[5]</sup>。典型的客观评估方法有:感知语音质量测量(PSQM)<sup>[6]</sup>、感知分析测量系统(PAMS/PSQM+)<sup>[7]</sup>、语音质量感知评估算法模型(PESQ)<sup>[8]</sup>等。但此类方法需通过比较原始输入信号和受损输出信号的失真才能评测出语音质量(这种方法也常被称为侵入式客观评价方法),因此无法进行语音质量在线评测。ITU-T G. 107 提出了 E-Model 模型,其

算法思想是将话音传输过程中若干因素对 QoE 的负面影响综合为参数  $R$ ,  $R$  值越大,表明话音质量越好,该模型不需要输入原始语音信号且具有较高的评估准确度<sup>[9]</sup>。但 E-Model 是一个以网络传输规划为目的的评价模型,在模型中甚至假设 IP 网络的损伤可以忽略<sup>[10]</sup>;即使在 E-Model 的改进版本中,也仅能研究 G. 711 和 G. 729 编码在指定分组大小下的语音质量<sup>[11]</sup>。间接测量模型主要以 Gap-Model<sup>[12]</sup>等为代表,其核心思想是通过实时监测 IP 分组的信息去估计在 IP 网上传输语音的收听质量,但 Gap-Model 模型缺乏对 VoIP 系统特性的分析,模型建立缺乏严格的理论与实验证明,本质上还是属于对主观评测方法的进一步扩展,也难以应用到模型设定之外的其它语音编码。

VoIP(Voice over Internet Protocol)是采用分组交换技术实现语音业务传输的一种新兴通信服务,与传统电话业务不同,IP 网络 QoS(分组丢包、延迟和抖动)的变化将会对 QoE 带来新形式的质量降级。为此,本文通过对 VoIP 系统各部

到稿日期:2010-06-29 返修日期:2010-10-01 本文受国家自然科学基金(61073154)资助。

胡治国(1977—),男,博士生,主要研究方向为网络性能评估, E-mail: huliang393@yahoo.com.cn; 张大陆(1949—),男,教授,博士生导师,主要研究方向为计算机网络、P2P 技术、语义 Web 等, E-mail: daluz@acm.org(通信作者); 张俊生(1984—),男,硕士生,主要研究方向为网络性能测量。

分组成的研究,用 Gilbert 模型对网络分组丢失特性进行了模拟,在运用 PESQ 算法对分组丢失损伤的结果进行分析的基础上,对 E-Model 模型的损伤系数:  $I_{e\_eff}$  进行了扩展,此方法使 E-mode 模型可方便地应用于其它语音编码,另一独特之处是通过分析 IP 网络 QoS 变化对语音 MOS 值的影响范围,根据实验数据建立了它们之间的映射模型。实验表明该模型可准确评价语音 QoE,又便于嵌入到网络监测系统中对语音 QoE 状况进行实时监测,有针对性地分析语音 QoE 降级原因,优化网络性能。

## 2 VoIP 系统数据传输特性分析

### 2.1 VoIP 系统数据传输框架

VoIP 系统(不包括信令部分)由发送端、IP 网络、接收端 3 部分组成。发送端采集语音并进行量化、编码,然后分组打包发送至 IP 网络,接收端收到数据包后提取出内部的有效载荷部分并重新组装成语音帧数据,经过抖动缓存平滑抖动后交付解码器解码,还原出语音信息。在 VoIP 系统中,当网络发生拥塞时,网络结点来不及转发同时到达的过多分组,或 IP 出现比特差错,包含语音帧数据的网络分组将会被丢弃,由于 FIFO 的“丢尾”特性,分组丢失具有突发特点。VoIP 系统端到端延迟可以分成两个部分:固定延迟和可变延迟。固定延迟包括编解码器引入的延迟和打包延迟,它与采用的压缩算法、打包的语音数据量相关<sup>[13]</sup>,由设备本身特性决定,它不随网络运行情况变化而变化。可变延迟则随网络运行状况的改变而动态变化,包括承载网上的传输、节点中排队、服务处理延迟和去抖动延迟。可变延迟与设备的端口速率、网络的负载情况和经过的网络路径等密切相关。整个 VoIP 系统各组成部分对语音质量的影响作用如图 1 所示。

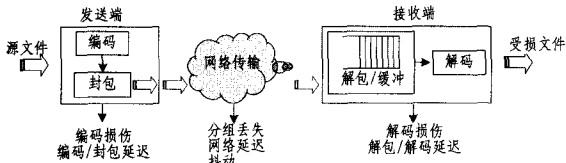


图 1 VoIP 系统数据传输框架<sup>[14]</sup>

### 2.2 IP 网络分组丢失特性

IP 网络中分组丢失之间往往存在短暂的相关性,分组的丢失与否是个二值事件。用随机变量  $X_i$  代表分组的丢失事件,  $X_i=0$  表示分组丢失,而  $X_i=1$  表示分组未丢失。则第  $i$  个分组丢失的概率为  $p = [X_i | X_{i-1}, X_{i-2}, \dots, X_{i-n}]$ ,  $X_{i-1}, X_{i-2}, \dots, X_{i-n}$  取所有的组合情况。当  $N=2$  时,该 Markov 链退化为著名的 Gilbert 模型<sup>[15,16]</sup>,可用两个独立的事件概率来完整地描述。

设状态  $X_i=0$  表示接收正常,  $X_i=1$  表示数据包丢失,那么当前状态  $X_i$  取值仅依赖于前一状态  $X_{i-1}$  的值。令状态转移概率  $p$  表示由“0”状态转变到“1”状态的概率,又可表示为  $P_{01}$ ,  $q$  表示由“1”状态转变到“0”状态的概率,又可表示为  $P_{10}$ 。  $1-p$  表示为  $P_{00}$ ,即  $P(X=0|X=0)$ ,  $1-q$  表示为  $P_{11}$ ,即  $P(X=1|X=1)$  状态转化如图 2 所示。

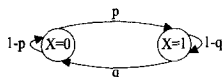


图 2 Gilbert 丢包模型

设  $P_1$  为总平均丢包率,  $P_0$  为总平均不丢包率,  $P_1$  记为 ULP(Unconditional loss probability, 无条件丢失概率),  $P_{11}$  记为 CLP(Conditional loss probability, 条件丢失概率)。  $p$  和  $q$  可以分别用 ULP 和 CLP 来表示<sup>[15,16]</sup>。

$$q = 1 - CLP, p = \frac{ULP(1 - CLP)}{1 - ULP} \quad (1)$$

实际上 ULP 即为网络的平均丢包概率,而 CLP 则反映网络丢包间的相关性。如果  $p+q=1$ , Gilbert 模型则转化为只有一个状态的 Bernoulli 模型。 Gilbert 模型描述了连续丢包长度  $k>0$  概率的几何分布,设随机变量  $Y$  用来描述突发丢包事件中连续丢包长度为  $k$  的概率<sup>[15]</sup>,即:

$$P(Y=k) = P(X=1|X=1)^{k-1} P(X=0|X=1) = (1-q)^{k-1} q \quad (2)$$

则  $Y$  的期望就是一次突发丢包事件的平均丢包长度<sup>[15]</sup>:

$$E[Y] = \sum_{k=0}^{\infty} k P(Y=k) = (1-CLP) \sum_{k=1}^{\infty} k CLP^{k-1} = \frac{1}{1-clp} = \frac{1}{q} \quad (3)$$

与随机丢包模型相比, Gilbert 模型采用无条件丢失概率和条件丢失概率对网络分组丢失行为进行建模,它可较好地模拟网络中分组丢失的突发特性<sup>[15]</sup>。

## 3 语音 QoE 评价及实验结果

### 3.1 语音质量评价算法

#### 3.1.1 PESQ 算法

PESQ 算法思想:对原始信号(参考信号)和通过测试的信号调整到标准听觉电平,再用输入滤波器模拟标准电话听筒进行滤波。对通过电平调整和滤波后的两个信号在时间上对准,并进行听觉变换,这个变换包括对系统中线性滤波和增益变化的补偿和均衡。两个听觉变换后的信号之间差异作为扰动(即差值)。分析扰动曲面提取出两个失真参数,在频率和时间上累积起来,映射到对主观平均意见分的预测值, PESQ 能够针对编解码和误码失真、随机或突发分组丢失情况给出精确的预测值<sup>[8]</sup>。

用 PESQ 预测语音质量只用到两个参数的组合:一个对称干扰 dSYM 和一个非对称干扰 dASYM,通过式(4)给出 PESQ 中映射得分:

$$PESQ_{mos} = 4.5 - 0.1dSYM - 0.0309dASYM \quad (4)$$

PESQ 原始得分范围为  $-0.5 \sim 4.5$ ,其扩展建议书中包含映射函数:

$$y = 0.999 + 4 / (1 + e^{-1.3669x + 3.8224}) \quad (5)$$

可以将得分转化为 ITU-T 800 的 MOS 评分<sup>[2]</sup>,分值范围为  $1.04 \sim 4.64$ ,其中  $x$  为 PESQ 测出的原始值。

#### 3.1.2 E-Model 算法

E-Model 基本原理:将语音质量损伤因素与人类心理上的感知特性的关系通过数学模型对应起来,用客观模型来模拟主观感觉的评价, E-Model 以  $R$  因子表示,也就是说  $R$  值是一种对传输品质的一个计量分数,用于度量端到端的语音质量, E-Model 的  $R$  参数由式(6)确定:

$$R = Ro - Is - Id - I_{e\_eff} + A \quad (6)$$

$Ro$ : 无网络延时和设备损伤因素的基本信号与收发噪声以及电流、背景噪声之比,即基本信噪比;  $Is$ : 表示语音传输时所产生的同步损伤,如由量化、连接和回音产生的影响;  $Id$ : 表示由语音信号所造成的延迟损伤,包括由于通话回音和交互

性丧失带来的影响;  $I_{e,eff}$ :有效设备损伤系数,代表由低比特率编解码产生的损伤。它也可包含由于信息包丢失带来的损伤;  $A$ :表示优势损伤因素,对于传统电话  $A$  取值常为  $0^{[16]}$ 。

$R$  值和用户满意度之间的关系如图 3 所示。

$R$	用户满意度	MOS	
100	很满意	4.5	理想区间
94, 3		4.4	
90	满意	4.3	
80		4.0	
70	有些用户不满意	3.6	可接受区间
60	许多用户不满意	3.1	不可接受区间
50	几乎所有用户不满意	2.6	
0	(没有定义)	1	

图 3  $R$  值和用户满意度之间关系<sup>[11]</sup>

E-Model 综合参数  $R$  与 MOS 值转化关系公式如下<sup>[10]</sup>:

$$MOS = \begin{cases} 1, & R < 0 \\ 1 + 0.035R + R(R-60)(100-R)7 \cdot 10^{-6}, & 0 < R < 100 \\ 4.5, & R > 100 \end{cases} \quad (7)$$

### 3.2 分组丢失对语音 QoE 损伤的影响

以 Gilbert 分组丢失模型为参考,通过在发送端主动控制数据包的发送模式,来控制传输语音流的平均分组丢失率  $ULP$  与突发分组丢失  $CLP$  取值,  $ULP$  取值范围:  $0\% \sim 5\%$ <sup>[17]</sup>,在  $0\% \sim 4\%$  之间,步长为  $0.1\%$ ;在  $4\% \sim 5\%$  之间,步长为  $0.5\%$ 。  $CLP$  取值范围为  $0.3 \sim 0.8$ <sup>[18]</sup>,步长为  $0.1$ 。实验中的编解码采用 ITU-T 提供源代码<sup>[19]</sup>,源语音数据采用 ITU-T P. 50 提供参考文件<sup>[20]</sup>,依据文献<sup>[13]</sup>设定分组间隔。语音 QoE 质量评价模型使用 ITU-T P. 862 (PESQ) 标准,实验框架如图 4 所示。

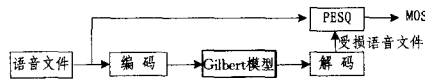


图 4 实验框架

实验中发现丢包位置不同, PESQ 取值会略有不同,因此需进行多次实验求平均值以消除实验过程中的随机误差。本节所有编码实验数据取 20 次实验的平均值,结果如图 5 所示。

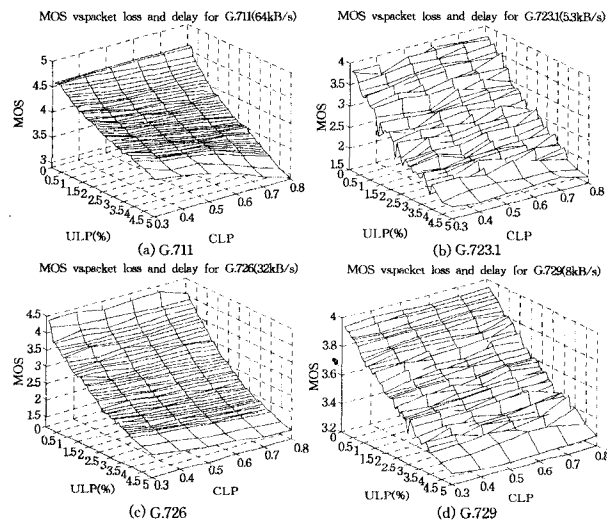


图 5 分组丢失与 MOS 的关系曲线

从实验结果可得以下结论:随着总平均丢包率  $ULP$  的逐

渐上升, MOS 值呈现明显下降趋势;在  $ULP$  相同的情况下,  $CLP$  值的增加也会使 MOS 值降低;以 MOS 大于等于 3.6 为阈值(多数用户感到满意)<sup>[2]</sup>, G. 711 在  $ULP = 5.5\%$ ,  $CLP = 0.3$ ;  $ULP = 5\%$ ,  $CLP = 0.4$ ;  $ULP = 4.5\%$ ,  $CLP = 0.5$ ;  $ULP = 3.8\%$ ,  $CLP = 0.6$ ;  $ULP = 3.1\%$ ,  $CLP = 0.7$ ;  $ULP = 2.4\%$ ,  $CLP = 0.8$ ; G. 723. 1 在  $ULP = 0.5\%$ ,  $CLP = 0.3$ ;  $ULP = 0.5\%$ ,  $CLP = 0.4$ ;  $ULP = 0.4\%$ ,  $CLP = 0.5$ ;  $ULP = 0.4\%$ ,  $CLP = 0.6$ ;  $ULP = 0.3\%$ ,  $CLP = 0.7$ ;  $ULP = 0.2\%$ ,  $CLP = 0.8$ ; G. 726 在  $ULP = 0.9\%$ ,  $CLP = 0.3$ ;  $ULP = 0.7\%$ ,  $CLP = 0.4$ ;  $ULP = 0.6\%$ ,  $CLP = 0.5$ ;  $ULP = 0.6\%$ ,  $CLP = 0.6$ ;  $ULP = 0.6\%$ ,  $CLP = 0.7$ ;  $ULP = 0.3\%$ ,  $CLP = 0.8$ ; G. 729 编码  $ULP = 3\%$ ,  $CLP = 0.3$ ;  $ULP = 3\%$ ,  $CLP = 0.4$ ;  $ULP = 2.4\%$ ,  $CLP = 0.5$ ;  $ULP = 2.4\%$ ,  $CLP = 0.6$ ;  $ULP = 0.3\%$ ,  $CLP = 0.7$ ;  $ULP = 2.4\%$ ,  $CLP = 0.8$  的情况下为阈值临界点。

### 3.3 单向延迟对语音 QoE 损伤的影响

分组延迟,特别是单向延迟是影响 VoIP 会话质量的重要因素<sup>[14,21]</sup>。其延迟由算法延迟、编/解码延迟、传输延迟和缓存延迟 4 部分组成。算法延迟:语音处理以语音分组为单位,当音频数据采样量达到一定数量后才进行编码,基于目前的编码解码方式,这种延迟无法避免。编/解码处理延迟:语音分组进行编码解码、回声消除、抖动算法、处理器间通信而引入的延迟。此外,许多使用模块处理的编码器还具有预测 (look-ahead) 功能,该功能在模块编码之前,需要对后来的语音样本进行缓冲,这增加了算法的延迟值。传输延迟:由于网络物理、路由等原因引起的延迟,是 VoIP 系统延迟的最主要部分。抖动缓冲延迟:语音分组进入接收端后,直到被抖动缓冲控制算法决定交给解码器解码播放之前的延迟<sup>[13]</sup>。

为分析网络传输过程中分组延迟与语音质量之间关系,将 E-Model 模型中的其他参数设置为 ITU-TG. 107 中推荐使用的默认值,化简后得:

$$R = 93.2 - I_d - I_{e,eff} \quad (8)$$

由于设分组丢失率为零,  $I_{e,eff}$  仅表示编码器对质量的损伤,其取值如表 2 所列。

表 2 编码引入延迟值<sup>[23]</sup>

编码类型	帧大小 (ms)	Look-ahead (ms)	$I_{e,eff}$	编码引入延迟 (ms)
G. 711 (64Kb/s)	0.125	0	0	0.375
G. 726 (32 kb/s)	0.125	0	7	0.375
G. 729 (8 kb/s)	10	5	10	35
G. 723.1 (5.3 Kb/s)	30	7.5	19	97.5

$I_d$  计算公式如下:

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3)^{[23]} \quad (9)$$

式中,  $d$  为单向延迟取值,  $d < 177.3$  ms,  $H(x) = 0$ , 否则  $H(x) = 1$ ;其转换曲线如图 6 所示。依据 E-Model 模型中  $R$  与 MOS 转换式(7)得到单向延迟取值与 MOS 之间的对应关系,如图 7 所示。

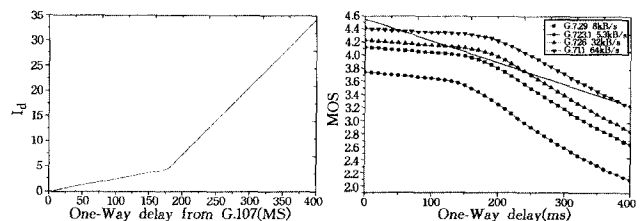


图 6 延迟与延迟损伤  $I_d$  转换

图 7 单向延迟与 MOS 评分的关系曲线

从实验结果中得出以下结论:在分组丢失为0的情况下,G711在延迟值320ms处;G.726在延迟值260ms处;G729在延迟值230ms处;G723.1在延迟值130ms处,MOS值约为3.6,排除语音编/解码带来的MOS值损伤影响,则网络传输延迟G711应控制在320ms以下;G.726应控制在260ms以下;G729应控制在295ms以下;G723.1应控制在32.5ms以下才可以得到满意的语音质量。

### 3.4 分组丢失、延迟综合作用对语音 QoE 损伤的影响

分组丢失、延迟与抖动是传输网络性能的主要指标,在现实网络环境中,它们总是相互影响、密不可分。另一方面,网络中数据包的抖动由于端系统抖动缓存的作用又转化为延迟<sup>[14,22]</sup>。因而,研究分组丢失率和延迟综合作用对语音 QoE 的影响有着更现实的意义。

为研究分组丢失、延迟综合作用对 QoE 损伤的影响,本文综合 2.2 节及 2.3 节的实验结果对 E-Model 进行了扩展,算法设计如下:

Step 1 用 PESQ 算法研究 Gilbert 丢包模型对语音质量的影响,将 PESQ 得分用式(5)转换为 MOS 值;

Step 2 将 Step1 中得到的 MOS 值依据式(10)转化成 E-Model 中的 R 值<sup>[12]</sup>;

在范围  $6.5 \leq R \leq 100$ , 能用下列公式从 MOS 计算 R:

$$R = \frac{20}{3} (8 - \sqrt{226} (h + \frac{\pi}{3})) \quad (10)$$

式中,

$$h = \frac{1}{3} \arctan 2 [18566 - 6750MOS,$$

$$15\sqrt{-903522 + 1113960MOS - 202500MOS^2}]$$

$$a \tan 2(x, y) = \begin{cases} \arctan(\frac{y}{x}), & \text{for } x \geq 0 \\ \pi - \arctan(\frac{y}{-x}), & \text{for } x < 0 \end{cases}$$

Step 3 在仅考虑分组丢失情况下,简化 E-Model 公式为:

$$I_{e,eff} = R_0 - R \quad (11)$$

对分组丢失实验所得数据,运用多元回归的方法建立 G.711, G.726, G.729, G.723.1 编码在不同分组丢失率 ULP, CLP 组合下与  $I_{e,eff}$  的映射关系。

$$I_{e,eff} = a + b * ULP + c * CLP + d * ULP^2 + e * CLP^2 + f * ULP * CLP \quad (12)$$

回归方程参数取值如表 3 所列。

表 3 回归方程的参数取值与检验

Standard	a	b	c	d	e	f	R2
G.711 (64Kb/s)	-9.96	885.49	19.29	-10585.94	-7.53	389.92	0.981
G.726 (32 kb/s)	4.69	1693.64	17.17	-17638.63	-9.82	306.62	0.982
G.729 (8 kb/s)	14.24	437.72	2.44	-2164.25	-1.56	-22.99	0.976
G.723.1 (5.3 Kb/s)	18.04	1453.51	-1.22	-13069.93	-0.29	60.54	0.973

4 种编码的标准偏  $\sigma$  分别为:1.219, 1.917, 0.652, 1.940; 回归方程显著性 F 检验分别为 2567.1, 2641.93, 2034.20, 1760.45。

方程的检验主要是通过式(14)标准偏差  $\sigma$  检验、式(16)

拟合优度  $R^2$  检验、式(17)回归方程显著性 F 检验进行评价。从方程检验结果来看,方程拟合程度优良,回归效果显著。

需要说明的是:虽然文献[11]也定义了  $I_{e,eff}$  计算公式:

$$I_{e,eff} = I_e + (95 - I_e) * \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl} \quad (13)$$

BurstR 是所谓的突发比<sup>[11]</sup>, Ppl 为分组丢失率(Gilbert 模型中的 ULP), 信息包丢失强度系数 Bpl 定义为编解码的特定值<sup>[24,25]</sup>, 但文献[11, 24, 25]提供的仅是 G.711, G.729 在特定分组大小情况下的数值。

Step 4 依据式(9)计算  $I_d$  值;

Step 5 简化 E-Model 的 R 参数计算式为:

$$R = R_0 - I_d - I_{e,eff} \quad (14)$$

Step 6 将 E-Model 综合参数 R 转化 MOS 值。

依据以上算法进行实验,得到 CLP=0.5 时 G.711 (64kbps), G.726(32kbps), G.729(8kbps) 和 G.723.1 (5.3kbps) 编码以及延迟与分组丢失情况对 MOS 评分的影响情况,如图 8 所示。

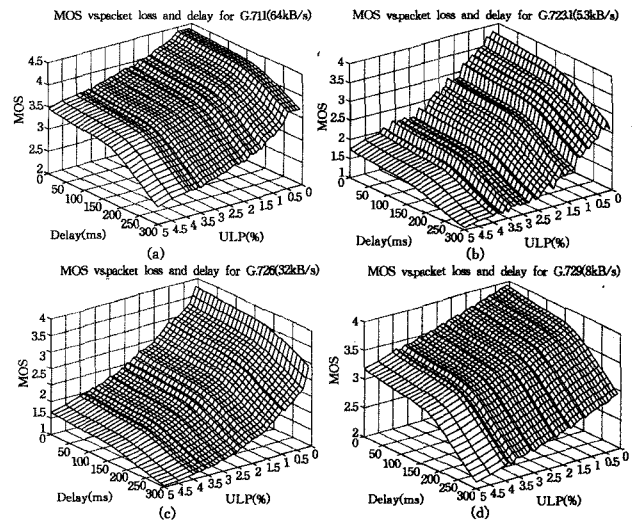


图 8 分组丢失和延迟对语音质量的影响

从实验结果中,得出以下结论:G.711 编码的语音质量最好,在 CLP 为 0.3, 延迟为 200ms, ULP 在 0~5% 的情况下 MOS 值都可达到 3.6; 根据文献[17]设定延迟阈值为 150ms 的情况下, G.711 编码在 CLP=0.4, ULP=3.4; CLP=0.4, ULP=3.4; CLP=0.5, ULP=2.9; CLP=0.6, ULP=2.3; CLP=0.7, ULP=2; CLP=0.8, ULP=1.7, G.726 编码在 CLP=0.3, ULP=0.5; CLP=0.4, ULP=0.4; CLP=0.6, ULP=0.4; CLP=0.7, ULP=0.2; CLP=0.8, ULP=0.2, G.729 编码在 ULP=1.2, G.723.1 编码在 ULP=0.4 的情况下 MOS 值为 3.6 以上。

## 4 模型建立与验证

为了更直接、更准确地分析 QoS 参数与语音 QoE 的映射关系,采用多项式回归的方法分析它们之间的联系。这是因为根据数学上级数展开的原理,任何曲线、曲面、超曲面的问题,在一定的范围内都能够用多项式任意逼近。所以当因变量与自变量之间的确实关系未知时,可用适当幂次的多项式来近似反映。

#### 4.1 模型性能指标及模型建立

对方程的检验主要是标准偏差、拟合优度检验(判定系数  $R^2$  的检验)和回归方程的显著性检验( $F$  检验)。其中标准偏差  $\sigma$  用来衡量回归方程评测数据与实测数据的偏离程度,值越小,反映模型误差越小,客观评价测度的性能越好;拟合优度  $R^2$  检验用来评价模型的拟合程度,  $R^2$  越接近 1,模型的拟合程度越高;显著性  $F$  检验整个回归系数是否有意义。在检验公式中,  $\hat{y}$  样本回归拟合得到 MOS 值,  $y_i$  代表实测的 MOS 值,  $\bar{y}$  代表实测 MOS 的平均值。

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{y}]^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

平方和分解公式:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y})^2 + \sum_{i=1}^n (\hat{y} - \bar{y})^2$$

记为:  $L_{yy} = Q + U$

$$F = \frac{U/m}{Q/(n-m-1)} \sim F(m, n-m-1) \quad (17)$$

拒绝域  $\chi_0 = \{F > F_\alpha(m, n-m-1)\}$ , 则认为回归效果好,  $m$  为自变量的个数,  $n$  为观测组数,  $\alpha$  为给定的置信度, 本文中  $m=3, n=7812, \alpha$  设为: 0.10, 0.05, 0.01。由  $F$  分布表可查得拟合方程所得  $F$  值都远大于  $F_\alpha(m, n-m-1)$ , 即回归方程总体回归效果显著。

本文使用数据拟合软件 Datafit<sup>[26]</sup> 完成多元回归方程的建立, 在其计算所得的多种形式的回归方程中, 以高的拟合优度作为方程形式选择的优先条件, 同时要求方程形式简单, 并避免加法及乘法以外的计算方式, 这是为了便于计算机移位操作, 加快方程运算速度, 从而易于在实际应用中使用。

G. 711, G. 726, G. 729, G. 723. 1 编码多元回归方程形式见式(18), 参数具体取值如表 4 所列。

$$MOS = a * Delay + b * ULP + c * CLP + d \quad (18)$$

式中, 延迟值 Delay 单位为 ms, ULP 取值 0%~5%<sup>[17]</sup>, CLP 取值 0.3~0.8<sup>[18]</sup>。

表 4 回归方程的参数取值与检验

Standard	a	b	c	d	$\sigma$	$R^2$	F
G. 711 (64kb/s)	-0.003	-26.64	-0.887	5.11	0.147	0.9045	24645.9
G. 726 (32kb/s)	-0.003	-51.05	-0.599	4.295	0.206	0.9215	30555.6
G. 729 (8kb/s)	-0.003	-16.14	-0.009	4.111	0.124	0.8903	21127.1
G. 723.1 (5.3kb/s)	-0.003	-43.65	0.014	3.744	0.188	0.9147	27952.3

#### 4.2 模型对比验证

为验证映射模型正确性, 选用在建模过程中未涉及的 QoS 参数组合: ULP=4.3%, ULP=4.8%(G. 711 编码)代入到 E-Model 中计算得到的 MOS 值与代入式(18)中得到的 MOS 值进行对比。实验结果: 拟合优度检验  $R^2$  分别为: 0.8923, 0.8846, 绝大多数误差在  $\pm 0.2$  之间, 证明了模型的准确性。误差曲面如图 9 所示。实验中使用 IBMR60i 笔记本, 用 matlab7.0b 实现相关算法, 本文提出的映射模型式(18)运算时间为 0.000005 秒, 满足网络监测实时性的要求。

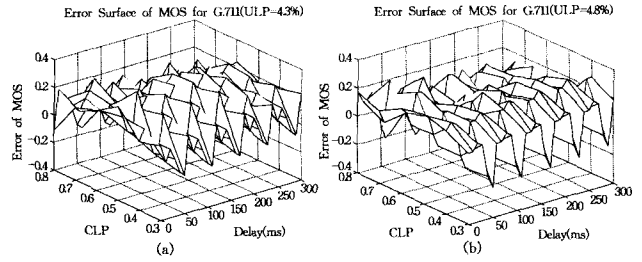


图 9 误差曲面(a)ULP=4.3%, (b)ULP=4.8%

**结束语** 本文围绕 QoS 与 VoIP 语音 QoE 的相关性展开, 研究 QoS 参数变化与 QoE 之间的映射关系, 用回归分析的方法建立了可测量 QoS 参数与 QoE 之间的映射模型, 其中, 分析了单个 QoS 参数对 QoE 的影响阈值, 并通过对 E-Model 算法的扩展, 研究了 QoS 参数综合作用情况下 QoE 的变化情况。在实验过程中, 多次重复实验保证了数据的可靠性, 验证实验也证明了评价模型的有效性, 与其它研究相比, 我们的研究更贴近网络实际, 映射模型形式更简单实用, 因此网络运营商和普通用户都可借助本文模型去预测语音 QoE 状况, 分析网络性能瓶颈。在这些研究成果的基础上进一步研究相应网络性能参数的测量算法, 最终建立以 QoE 为评价标准的网络监测体系将是下一步研究工作的重点。

#### 参考文献

- [1] Rahrer T, Fiandra R, Wright S. Triple-play Services Quality of Experience(QoE) Requirements and Mechanisms[M]. DSL Forum Working Text WT-126, February 2006
- [2] ITU-T, Recommendation P. 800-Methods for subjective determination of transmission quality[S]. International Telecommunication Union-Telecommunication Standardisation Sector, August 1996
- [3] Voiers William D. Diagnostic Evaluation of Speech Intelligibility [M] // M. E. Hawley, ed. Benchmark Papers in Acoustics. Stroudsburg, PA; Dowden, Hutchenson and Ross, 1977: 374-387
- [4] ITU-T, Recommendation P. 830-Subjective performance assessment of telephone-band and wideband digital codecs[S]. International Telecommunication Union-Telecommunication Standardisation Sector, February 1996
- [5] 陈国, 胡修林, 张蕴玉, 等. 语音质量客观评价方法研究进展[J]. 电子学报, 2001, 29(4): 548-552
- [6] ITU-T, Recommendation P. 861-Objective quality measurement of telephone-band speech codecs[S]. International Telecommunication Union-Telecommunication Standardisation Sector, February 1998
- [7] Rix A, Hollier M. The perceptual analysis measurement system for robust end-to-end speech assessment[A] // Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing [C]. Istanbul, Turkey, IEEE Press, June 2000(3): 1515-1518
- [8] ITU-T, Recommendation P. 862-Perceptual Evaluation of Speech Quality(PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs[S]. International Telecommunication Union-Telecommunication Standardisation Sector, February 2001
- [9] ITU-T, Recommendation G. 107-The E-Model, A computational model for use in transmission planning[S]. International Telecommunication Union-Telecommunication Standardisation Sector, March 2000

(下转第 70 页)

- Filter in Multitarget Tracking[C]//Proceedings of the Fifth Symposium on Non-linear Estimation. 1974;16-21
- [2] Bar-Shalom Y, Tse E. Tracking in a cluttered environment with probabilistic data association[J]. *Automatica*, 1975, 11(9): 451-460
- [3] Turkmen I, Guney K. Cheap Joint Probabilistic Data Association with Adaptive Neuro-Fuzzy Inference System State Filter for Tracking Multiple Targets in Cluttered Environment[J]. *International Journal of Electronics and Communications*, 2004; 349-357
- [4] Roecker J A, Phillis G L. Suboptimal Joint Probabilistic Data Association [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1993, 29(2): 510-517
- [5] Fitzgerald R J. Development of practical PDA logic for multitarget tracking by microprocessor [M]. *Multitarget-Multisensor Tracking: Advanced Applications*. 1990; 220-232
- [6] Roecker J A, Phillis G L. A Class of Near Optimal JPDA Algorithms[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1994, 30(2): 504-510
- [7] 秦卫华, 胡飞, 秦超英. 一种简化的联合概率数据关联算法[J]. *西北工业大学学报*, 2005, 23(2): 276-279
- [8] 邓未央, 王宝树. 一种近似的联合概率数据互联算法[J]. *电光与控制*, 2004, 11(2): 42-44
- [9] Zhou B, Bose N K. Multitarget Tracking in Clutter; Fast Algorithms for Data Association[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1993, 29(2): 352-363
- [10] 朱大奇, 史慧. *人工神经网络原理及应用*[M]. 北京: 科学出版社, 2006; 107-110
- [11] 陈小惠, 王建华, 黄国建. 基于 Kohonen 神经网络的多传感器多目标跟踪方法[J]. *中国造船*, 2000, 41(4): 66-70
- [12] 刘以安, 曹奇英, 刘同明, 等. 基于遗传模拟退火算法的机动多目标数据关联问题研究[J]. *华东船舶工业学院学报*, 2000, 14(6): 32-37
- [13] 赵智勇. 机动目标跟踪理论的算法研究[D]. 无锡: 江南大学, 2008
- [14] 敬忠良. *神经网络跟踪理论及应用*[M]. 北京: 国防工业出版社, 1995; 58-60
- [15] Singer R A, Stein J J. An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems[C]//Proceedings of the 1971 IEEE Conference on Decision and Control. Miami Beach, 1971; 171-175
- [16] Blom H A P. A Proceedings of the Efficient Filter for Abruptly Changing Systems[C]//Proceedings of the 13<sup>th</sup> IEEE Conference on Decision and Control. 1984; 656-658
- [17] Reid D B. An Algorithm for Tracking Multiple Targets[C]//Proceedings of the 17<sup>th</sup> IEEE Conference on Decision and Control. 1978; 1202-1211
- [18] Singer R A, Sea R G, Housewright K B. Derivation and evaluation of improved tracking filters for use in dense multi-target environments[J]. *IEEE Trans. Information Theory*, 1974, 20(7): 423-432
- [19] Sittler R W. An optimal data association problem in surveillance theory[J]. *IEEE Trans. Military Electronics*, 1964, 8(2): 125-139
- [20] Avitzour D. A maximum likelihood approach to data association [J]. *IEEE Trans. Aerosp. Electron. Syst*, 1992, 28(2): 560-565
- [21] Blom H A P, Bar-Shalom Y. The interacting multiple model algorithm for systems with markovian coefficients [J]. *IEEE Transactions on Automatic Control*, 1988, 33(8): 780-783
- [22] Mori S, Chong C Y, Tse E, et al. Tracking and classifying multiple targets without a priori identification[J]. *IEEE Trans. Automatic Control*, 1986, 31(5): 401-409
- [23] 赵开斌. 关于统计距离的一点注记[J]. *安庆师范学院学报*, 2002, 8(4): 1-2

(上接第 53 页)

- [10] 成卫青, 龚俭, 丁伟. 基于流特性和真值程度的 VoIP 语音质量单端客观评价[J]. *通信学报*, 2008, 29(4): 30-39
- [11] ITU-T, Recommendation G. 107-the E-model, a computational model for use in transmission planning[S]. *International Telecommunication Union-Telecommunication Standardisation Sector*. March 2005
- [12] Calyam P, Ekicio E, Lee C, et al. A gap-model based framework for online VVoIP QoE measurement[J]. *Journal of Communications and Networks*, 2007, 9(4): 446-56
- [13] ITU-T Recommendation G. 114-One-way transmission time[S]. *International Telecommunication Union Telecommunication Standardisation Sector*. May 2003
- [14] Markopoulou A P, Tobagi F A, Karam M J. Assessment of VoIP Quality over Internet Backbones[C]//IEEE INFOCOM 2002. New York, 2002; 150-159
- [15] 兰帆, 张尧弼. 基于 Gilbert 模型的网络丢包仿真[J]. *计算机工程*, 2004, 30: 200-202
- [16] Haßlinger G, Hohlfield O. The Gilbert-Elliott model for packet loss in real time services on the Internet[A]//14. GI/ITG Konferenz Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen (MMB) [C]. Dortmund, Germany, Mar. 2008; 269-283
- [17] ITU-T Recommendation G. 1010-End-user multimedia QoS categories[S]. *International Telecommunication Union Telecommunication Standardisation Sector*. October 2001
- [18] Uemura S, Fukumoto N, Yamada H, et al. QoS/QoE measurement system implemented on cellular phone for NGN [A]//Consumer Communications and Networking Conference, CCNC [C]. 2008; 117-121
- [19] Software tools for speech and audio coding standardization[EB/OL]. <http://www.itu.int/rec/T-REC-G.191-200509-I/en>
- [20] ITU-T Test signals for telecommunication systems [EB/OL]. <http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050>
- [21] Perlicki K. Simple analysis of the impact of packet loss and delay on voice transmission quality[J]. *Journal of telecommunications and information technology*, 2002; 53-56
- [22] Mohamed S, Rubino G, Varela M. A method for quantitative evaluation of audio quality over packet networks and its comparison with existing techniques[A]//Measurement of Speech and Audio Quality in Networks(MESAQIN)[C]. Prague, 2004
- [23] Cole R G, Rosenbluth J H. Voice over IP performance monitoring[J]. *Computer Communication Review*, 2001, 31(2): 9-24
- [24] ITU-T Recommendation G. 113(2002). Transmission impairments due to speech processing[S]. *International Telecommunication Union-Telecommunication Standardisation Sector*. May 2002
- [25] ITU-T Recommendation G. 113 Appendix I(2002). Provisional planning values for the equipment impairment factor  $I_e$  and packet-loss robustness factor  $B_{pl}$ [S]. *International Telecommunication Union-Telecommunication Standardization Sector*. May 2002
- [26] DataFit9.0 <http://www.oakdaleenr.com/>