

# 基于 E-Model 的 VoIP 语音质量研究

张泽明, 胡 磊, 任久春, 高传善

(复旦大学信息科学与工程学院复旦-华夏信息技术联合实验室, 上海 200433)

**摘 要:** 针对目前网络电话语音质量难以准确评价及预测的情况, 基于 E-Model 对 VoIP 的语音质量进行预测。分析几个主要影响因素, 如延时、丢包等对语音质量的影响, 构建 VoIP 语音质量预测模型, 将 E-Model 中未考虑到的抖动因素引入模型公式, 着重考虑抖动缓冲区的大小对语音质量的影响。通过设计相关验证实验, 证明该模型对 VoIP 语音质量的预测具有较高的准确度。

**关键词:** 互联网协议语音技术; 抖动缓冲; 语音质量

## Research on Speech Quality of VoIP Based on E-Model

ZHANG Ze-ming, HU Lei, REN Jiu-chun, GAO Chuan-shan

(Fudan-Huaxia Joint Laboratory on Information Technology, School of Information Science and Engineering, Fudan University, Shanghai 200433)

**【Abstract】** Aiming at the problem that how to evaluate and survey the speech quality of VoIP accurately, this paper evaluates the speech quality of VoIP based on E-Model, analyzes the impact of several major factors such as time delay, and packet loss. It constructs a model which includes the factor jitter that is not considered in E-Model and analyzes the impact of jitter buffer on speech quality. Results of the designed experiments show that the precision of the model evaluating the speech quality of VoIP is satisfied.

**【Key words】** Voice over Internet Protocol(VoIP); jitter buffer; speech quality

IP 电话凭借其巨大的价格优势得到广泛应用, 然而由于目前网络自身的原因, 不能保证其服务质量。对网络语音质量进行科学可靠的测量不仅是语音通信, 更是网络测量与网络规划设计中相当关键的问题。E-Model<sup>[1]</sup>作为预测语音质量的一种方法, 全面考虑了影响 IP 电话质量的多种因素, 如延迟、丢包、抖动等, 因此, 受到国内外研究人员的广泛关注。

### 1 E-Model 简介

E-Model 最早由欧洲的 ETSI 标准组织提出, 之后又由 ITU-T 标准化形成 G.107 建议。E-Model 是基于设备损伤的测量方法, 它关注数据全面的网络损伤因素, 可较好适应在数据网络中语音质量的评估。

整个通话系统被分为发送和接收 2 个部分, 电话两端的用户既是语音信息的发送者也是接收者。表 1 中的英文缩写作为计算模型的输入使用的传输参数。E-Model 中所有众多参数的计算都是由表 1 中的基本参数推导而来。

在 ITU-T P.800 中规定的平均主观值  $MOS$ , 基于该主观评测, 将人接听和感知语音质量的行为被调研和量化, 接听何种级别质量的语音, 就得到多少平均主观值  $MOS$ 。一般情况下,  $MOS$  值为 4 或更高则被认为是较好的语音质量, 若平均主观值  $MOS$  低于 3.6, 则多数接听者不能满意语音质量。平均主观测试虽然准确有效, 但该方法存在不易实现和高成本的问题。因此, 人们在不断地探索能进行客观测量的方法。E-Model 考虑语音信号传输过程中若干因素, 如延时、噪声、回音、编码器性能、丢包等网络损伤因素对语音质量的影响并将其综合为参数  $R$ , 用以评估该语音呼叫的主观品质。 $R$  越大, 表明语音品质越好, 其评价标准见表 2。

$R$  可通过如下公式转换成通用的评估分数  $MOS$  分值:

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

表 1 E-Model 参考连接基本参数

基本性能参数	缩写
发送响度等级	SLR
接收响度等级	RLR
侧音掩蔽等级	STMR
受话器侧音等级	LSSTR
发送侧电话 D 值	Ds
接收侧电话 D 值	Dr
扬声器回声强度等级	TELR
加权回声通道损耗	WEPL
回声通道的平均单向延时	T
4 线环往返行程延迟	Tr
无回声连接绝对延迟	Ta
量化失真单元数量	qdu
设备损伤系数	Ie
信息包丢失强度系数	Bpl
随机信息包丢失概率	Ppl
突发比	BurstR
以 0 dB 点为参考的电路噪声	Nc
接收噪声基底	Nfor
发送侧室内噪声	Ps
接收侧室内噪声	Pr
获益系数	A

**作者简介:** 张泽明(1984 - ), 男, 硕士, 主研方向: 计算机与网络通信; 胡 磊, 硕士; 任久春, 博士; 高传善, 教授、博士生导师

**收稿日期:** 2008-03-28 **E-mail:** cgao@fudan.edu.cn

表 2 R 值和 MOS 分值评价标准

R 值	用户满意度	MOS 分值
90~100	非常满意	4.3~5.0
80~90	满意	4.0~4.3
70~80	有些用户不满意	3.6~4.0
60~70	很多用户不满意	3.1~3.6
50~60	几乎所有用户都不满意	2.6~3.1
0~50	没有定义	1.0~2.6

E-Model 的基本原理如图 1 所示。

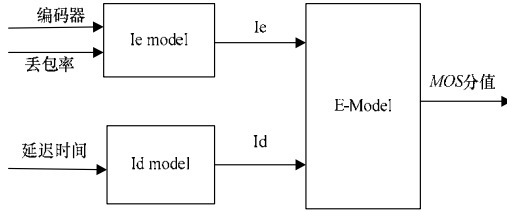


图 1 E-Model 基本原理

E-Model 的基本公式为

$$R = Ro - Is - Id - Ie_{eff} + A \quad (2)$$

$Ro$ : 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。

有效设备损伤系数  $Ie_{eff}$  代表由低比特率编解码产生的损伤。它也包含由于随机分布的信息包丢失带来的损伤。

$A$ : 表示优势损伤因素, 它是当有其他优势传递到用户的一种补偿, 对于传统电话  $A$  取值为 0。

式(2)中的参数因子多数与其他一些影响通话的因素有关, 并由较为复杂的公式求出。如基本信噪比  $Ro$  由式(3)计算而得:

$$Ro = 15 - 1.5(SLR + No) \quad (3)$$

其中的参数又由其他一些参数计算而得, 与响度评测值  $SLR$ ,  $RLR$ 、背景噪音  $Ps$ ,  $Pr$ 、电话  $D$  值  $Ds$ ,  $Dr$ , 扬声器回声强度等级  $TELR$ , 受话器侧音等级  $LSDR$  等因素有关。

$Is$ : 表示语音传输时所产生的同步损伤, 它可进一步细分为 3 个具体的损伤因素:

$$Is = Iolr + Ist + Iq \quad (4)$$

其中,  $Iolr$  表示由太低的  $OLR$  值所引起的质量下降;  $Ist$  表示的是由非最佳侧音引起的损伤;  $Iq$  表示量化失真所造成的损伤。影响  $Is$  的参数有: 响度评测值  $OLR$ ,  $RLR$ , 背景噪音  $Ps$ ,  $Pr$ , 电话  $D$  值  $Ds$ ,  $Dr$ , 说话者回音响度额定值  $TELR$ , 单向延迟时间  $T$ , 侧音屏蔽额定值  $STMR$  和量化失真度  $qdu$  等。

损伤系数  $Id$ : 代表由于声音信号延迟引起的所有损伤, 并再分为 3 个系数:  $Idte$ ,  $Idle$  和  $Idd$ :

$$Id = Idte + Idle + Idd \quad (5)$$

其中, 系数  $Idte$  是扬声器回声引起的损伤; 系数  $Idle$  代表由于受话器回声引起的; 系数  $Idd$  代表由于绝对延迟  $Ta$  太长带来的损伤, 即使使用良好的回声消除手段绝对延迟还是存在; 系数  $Id$  和单向延迟  $T$ 、绝对延迟  $Ta$ 、扬声器回声强度等级  $TELR$ 、侧音掩蔽等级  $STMR$ 、响度评测值  $OLR$ ,  $RLR$ 、背景噪音  $Ps$ ,  $Pr$ 、电话  $D$  值  $Ds$ ,  $Dr$ 、加权回声通道损耗  $WEPL$ 、4 线环往返行程延迟  $Tr$  等参数有关。

设备损伤系数  $Ie_{eff}$  计算公式为

$$Ie_{eff} = Ie + (95 - Ie) \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl} \quad (6)$$

其中,  $Ie$  为设备损伤系数默认值, 不同的编码方式对应不同  $Ie$  值, 在 I/G.113<sup>[2]</sup>附录中列出了不同编码方式对应的  $Ie$  值。

## 2 对 E-Model 的改进

应用 E-Model 给出的公式, 获取相关的参数, 就能预测 IP 电话的话音质量。但目前 E-Model 对 VoIP 语音的预测还有一定偏差, 如 E-Model 并未考虑到网络抖动对话音质量的影响, 而网络的抖动恰恰是影响 VoIP 语音质量的一个较为重要的因素, 本文对此进行进一步研究。

抖动也称作延时变化, 是指在一个 VoIP 呼叫过程中所有发送的数据包到达的时间差异程度。在理想状态下, 对话的每一个语音包传递到接收方的时间延迟应该是一致的, 但由于数据包在网络中会经由不同的路由传送到接收方, 就造成不同语音包之间的通路时间存在差异, 存在抖动的情况。抖动的大小取决于数据包的通路时间的差异程度, 差异越大, 抖动越大。在电话呼叫中, 它表现的效果与丢包产生的效果相似, 某些字词听不清楚或错误。因此, 对于评估语音质量来说, 抖动是必须要考虑的因素。

要消除抖动对话音的影响, 须加入抖动缓冲区。抖动缓冲区越大, 则防止抖动的效果越好, 但却增大了附加的延长时间, 因此, 抖动缓冲区的大小一定要合适。

现有的 E-Model 计算公式关注了延时对话音质量的影响, 却未引入对抖动的考虑, 因此, 最后得到的  $R$  分值偏高。为得到与实际更为符合的  $R$  值, 本文在 E-Model 中添加了抖动因子  $Ij$ <sup>[3]</sup>, 使式(2)式变为式(7):

$$R = Ro - Is - Id - Ie_{eff} - Ij + A \quad (7)$$

对于抖动因子  $Ij$ , 将其建模为式(8)

$$Ij = s_6 T^6 + s_5 T^5 + s_4 T^4 + s_3 T^3 + s_2 T^2 + s_1 T + s_0 \quad (8)$$

其中,  $T = \ln(1 + tj)$ ,  $tj$  为抖动缓冲区的大小, 单位为 ms;  $S_0 \sim S_8$  均为常数, 对于 G.711PCM 编码 ulaw, 常数取值为:  $S_6 = -0.1465$ ;  $S_5 = 3.3877$ ;  $S_4 = -30.9436$ ;  $S_3 = 139.0633$ ;  $S_2 = -305.0000$ ;  $S_1 = 257.2308$ ;  $S_0 = 24.766$ 。

## 3 实验模拟

本文搭建一个模拟的网络环境, 以验证改进的 E-Model 模型准确性。实验使用到 3 个开源软件: NIST Net, OpenPhone (OpenH323) 和 ITU-T PESQ v2.0。使用 OpenPhone 来发送和接收数据; 使用 ITU-T PESQ v2.0 来算出 MOS 分值; 在 Linux Red Hat 环境下用 NIST Net 模拟网络的情况, 并设置各种影响语音质量的参数。

实验中选用 PESQ 标准声音示例中的文件 a.wav (文件时长 8.4 s) 作为传送的语音数据, 并将其他大部分影响网络语音质量的参数如丢包率、延时时间等依照 E-Model 模型的默认值设置 (设置为 0)。Openphone 设置为: 编码方式为 G.711, ulaw, 采样频率 8 kHz, 码率 64 Kb/s, 帧长 30 ms。

分别进行 10 组实验, 每组实验只有抖动缓冲区  $jitter$   $buffer$  值在变化, 该值由 0 ms 一直变化到 400 ms (超过 400 ms 由于附加了过多的延时而没有实际使用价值, 故没有验证意义), 并将每个抖动缓冲容量对应的 10 组 MOS 数据取均值作为该缓冲容量对应的 MOS 分值, 并与添加了抖动因子  $Ij$  后的 E-Model 换算而得的 MOS 分值进行比较, 如图 2 所示, 当增加了抖动缓冲区长度时, MOS 分值上升很快, 当抖动缓冲区容量为 30 ms 时, MOS 分值上升到最高, 是一个令人比

较满意的话音质量,显示了抖动缓冲区对于话音质量的改善能力。当缓冲区逐渐增大时, MOS 分值又开始下降。当抖动缓冲区时长增至 100 ms 时及继续增大后, MOS 分值变化不大,趋于稳定。此时由于抖动缓冲附加的延时太大,导致话音质量同没有加入抖动缓冲区时相差不大,话音质量比较差,抖动缓冲区失去了改善话音的能力。

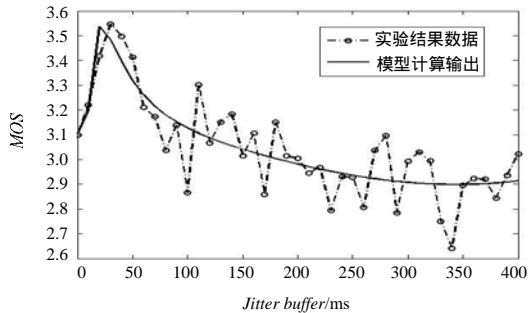


图2 实验结果

在实际应用中,一般核心路由器的抖动缓冲时延为 0 ms ~ 255 ms, Internet 的网间时延为 50 ms ~ 400 ms<sup>[4]</sup>,因此,本实验选定的抖动缓冲时延已经完全包含了实际中的应用范围。

延时的时长在 100 ms ~ 200 ms 之间开始被收听者所察觉,使会话不自然。建议的延时上限为 150 ms,若延时达到 200 ms 则开始有严重的会话中断<sup>[5]</sup>,本次实验显示,当抖动缓冲附加时延达到 100 ms 时,话音质量已经不算太理想,这与实际情况基本相符。

由于 MOS 分值本身带有一定主观性,因此该值相差在 0.5 范围之内,话音质量不会有太明显的感觉。由图 2 可见,添加了抖动因子的 E-Model 得到的 MOS 分值与实际实验数据能较好符合,最大的偏差不超过 0.3,实验得到的 MOS 分值

与实际听到的话音质量相符程度很高。

由于无法控制发送声音的背景噪声等参数,导致 MOS 分值相对较低,因此 E-Model 中参数  $R_0$  的值不能达到 E-Model 默认的最大值,本次实验中取值为 84.768 8,与实际比较相符。由于发送与接收端的时间差异,导致发送文件和接收端录音文件时长存在无法克服的时间偏差,而当抖动缓冲区的长度越长时,文件时长差异对于 PESQ 比较接收和发送文件计算而得的 MOS 分值的影响就越大,本次实验采取了多次实验取平均值的办法努力减小这种误差。由图 2 可知,当 jitter buffer 值大于 100 ms 后,实际曲线与理论曲线的误差在允许范围之内。

#### 4 结束语

E-Model 假定所有因子对于话音的影响都是加性影响的,本文将加性抖动因子引入 E-Model,对抖动缓冲区对话音的影响进行预测。实验结果显示,预测值与实际话音质量比较相符,实验结果较好。

#### 参考文献

- [1] ITU-T Recommendation. G.107-2005 The E-model, a Computational Model for Use in Transmission Planning[S]. 2005.
- [2] ITU-T Recommendation. G.113-2005 Transmission Impairments Due to Speech Processing[S]. 2001.
- [3] Ding Lijing, Goubran R A. Speech Quality Prediction in VoIP Using the Extended E-Model[C]//Proc. of Global Telecommunications Conference. San Francisco, USA: IEEE Press, 2003.
- [4] 柳 扬, 李 巍. VoIP 语音质量的测试[J]. 世界电信, 2002, 11(8): 28-31.
- [5] 段 洁. VoIP 语音质量测量新方法 E-Model[J]. 中国数据通信, 2004, 9(22): 63-67.

(上接第 277 页)

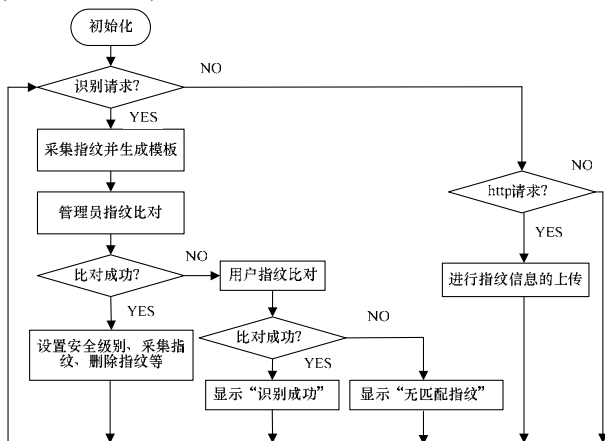


图5 总流程

#### 5 结束语

本终端利用了嵌入式系统的小型化、高性能、低价格等特点,完成对指纹的识别与管理;将 B/S 结构应用于嵌入式指纹平台,这样就可以通过 IP 地址实现对本终端的访问。本设计集自动指纹识别技术、数据存储技术、单片机嵌入式技术、计算机网络技术于一体,实现了对指纹终端的动态管理

和远程访问。

本设计的创新性在于:在一个嵌入式平台上同时完成了指纹信息的采集、识别、存储与管理,而且在其上架设了 WebServer。这样就无须配置服务器,将在服务器端进行的指纹存储与管理功能转移到了嵌入式平台。本终端充分利用了嵌入式系统支持操作系统及网络的特点,在保证功能的同时,大大节约了成本。本终端可应用于小型的门禁系统、考勤系统、指纹门锁等。

#### 参考文献

- [1] 骆 峰, 宁新宝, 尹义龙. 网络化指纹门禁系统设计[J]. 微计算机应用, 2003, 24(1): 23-26.
- [2] 李驹光. ARM 应用系统开发详解——基于 S3C4510B 的系统设计[M]. 北京: 清华大学出版社, 2004.
- [3] Gupta P, Ravi S, Raghunathan A. Efficient Fingerprint-based User Authentication for Embedded Systems[C]//Proc. of the 42nd Conference on Design Automation. New York, USA: IEEE Press, 2005: 244-247.
- [4] Yang Shenglin, Sakiyama K, Verbauwhede I. Efficient and Secure Fingerprint Verification for Embedded Devices[J]. EURASIP Journal on Applied Signal Processing, 2006, (3): 1-11.