

VoLTE 语音质量评价方法及测试方案

许 可 思博伦通信

编者按:截至2014年1月,全球已有超过40家运营商开始部署VoLTE服务。中国移动也宣布将在2014年底之前全网部署VoLTE服务。VoLTE和高清语音服务预期可以给客户提供更佳的语音用户体验,帮助运营商应对OTT语音冲击和ARPU值下降的不利趋势。对运营商而言,部署VoLTE将带来两方面的价值,一是提升无线频谱利用率、降低网络成本;二是提升用户体验。思博伦通信许可所撰《VoLTE语音质量评价方法及测试方案》一文分别对常用语音质量评价方法和Nomad-HD语音测试解决方案进行了深入分析和研究,可供相关技术人员参考。

1 引言

VoLTE即Voice over LTE,是基于LTE网络数据域的语音业务方案。该方案基于IMS,提供全IP通话。LTE网络是一种全IP网络,全部业务承载于数据域上,可实现数据与语音业务在同一网络下的统一。

截至2014年1月,全球已有超过40家运营商开始部署VoLTE服务。中国移动也宣布将在2014年底之前全网部署VoLTE服务。VoLTE和高清语音服务预期可以给客户提供更佳的语音用户体验,帮助运营商应对OTT语音冲击和ARPU值下降的不利趋势。

对运营商而言,部署VoLTE将带来两方面的价值,一是提升无线频谱利用率、降低网络成本;二是提升用户体验。

VoLTE的体验明显优于传统电路域语音。首先,高清语音和视频编解码的引入显著提高了通信质量;其次,VoLTE的呼叫接续时长大幅缩短,测试表明

VoLTE比CS呼叫缩短了一半以上。

VoLTE通过全IP的4G网络和IMS服务器提供语音服务,服务的部署需要网络侧和终端侧做出大量投资和研发。其实现原理和传统的2G、3G语音服务有很大不同,给运营商、终端厂商、芯片厂商带来了很大挑战。

如何判断重金投资的网络和研发的终端真正带来了更好的用户体验?如何衡量新的VoLTE语音服务语音质量优于2G、3G网络和几乎免费的OTT应用?如何衡量、评估多种操作系统多种终端在4G网络中语音性能孰优孰劣?

回答以上问题需要采用相应测试系统对VoLTE服务进行准确、标准、可重复以及自动化的专业测试评价。通常可以采用表1的KPI作为衡量VoLTE服务用户体验的指标。

其中,语音质量至关重要,本文第二、三章将分别

介绍常用语音质量评价方法和Nomad HD 语音测试解决方案。

表 1 KPI作为衡量 VoLTE 服务用户体验指标

用户体验	关键性能指标
呼叫可靠性	起呼成功率, 掉话率
起呼时间	呼叫建立时间
语音质量	MOS 分
延迟	口到耳语音延迟

2 语音质量评估方法介绍

语音质量的评估方法主要包括主观评价和客观评价两大类(见图1)。

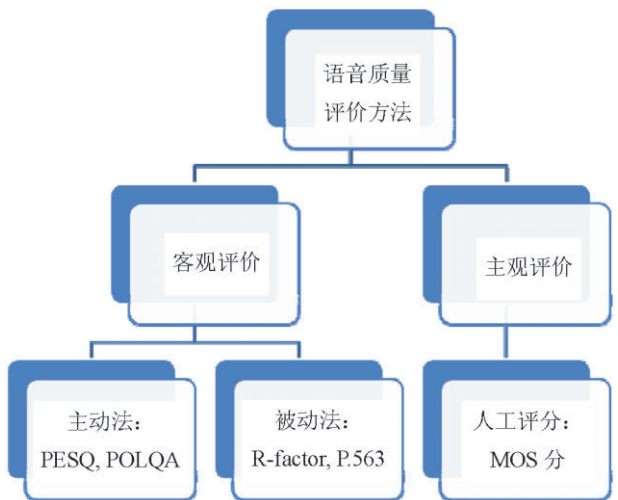


图 1 语音质量评估方法

主观评价指以人为主体的语音质量评价,由参与评价的评听人根据预先约定的评估准则对语音质量进行打分,它反映了评听人对语音质量好坏的一种主观印象。主观评价方法比较繁杂,为了排除偶然因素,减少评价波动方差,需要参与评价的评听人数量较多(一般40人以上)。但是,由于人是语音的最终接受者,这种评价方法是语音质量的真实反映。

客观评估是指用机器自动判别语音质量。它从原理上又可分为两类评价方式:基于输入/输出方式的主动式评估和基于输出方式的被动式评估。

主动式的评价是建立在原始语音信号和失真语音信号的误差对比上,大多数这种方式的客观评估采用数值距离或者描述听觉系统如何来感知质量的听觉模

型来量化语音质量的好坏;被动式的客观评价仅以语音系统的输出信号来评估语音质量的好坏。

(1)主观评价

国内外使用较多的主观评价方法包括平均意见分 MOS (Mean Opinion Score)法、韵母可懂度测量 DRT (Diagnostic Rhyme Tests)、失真平均意见分 DMOS (Degradation Mean Opinion Score)、满意度测量 DAM (Diagnostic Acceptability Measure)等。

在实际语音质量评价中,MOS法最为常见。ITU P.800标准定义了MOS评价方法。参加评价的评听人在听完测试语音后从表2的5个等级中选择某一级作为他对语音质量的评价,全体试验者的平均分就是所测语音质量的MOS值。

表 2 主观评价5个等级

级别	用户满意度
4.0~5.0	很好,听得清楚,延迟很小,交流流畅
3.5~4.0	稍差,听得清楚,延迟小,有点杂音
3.0~3.5	可以接受,有一定延迟,可以交流
1.5~3.0	勉强,听不太清,交流重复多次
0~1.5	极差,听不懂

主观性测试的优势是符合人对语音的感受;缺点是耗时、昂贵、可重复性低,受人的主观影响较大。

(2)主动式评价

主动式评价方法主要由ITU标准组织定义,评价方法的发展沿革历史如图2所示。

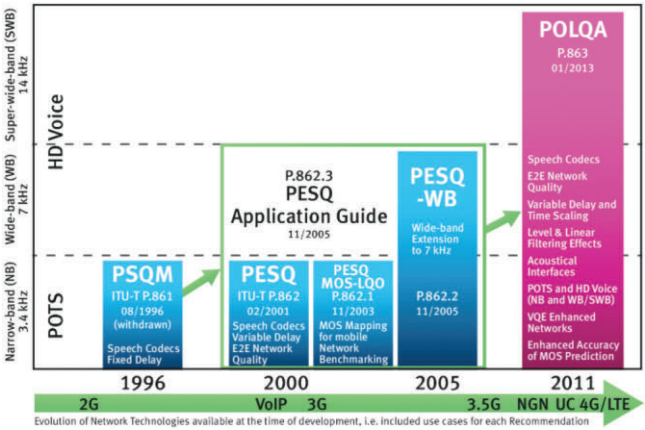


图 2 ITU组织感知语音评价方法发展历史

其中,PESQ和POLQA是目前仍然广泛使用的语音质量评价方法。PSQM由于种种缺陷目前在实际中

已经很少采用。

PESQ (Perceptual Evaluation of Speech Quality) 由英国电信和KPN共同开发,并在2001年被ITU采纳为P.862规范。它比较声源信号和退化信号并给出一个类似人工听力评估测试的MOS分值。PESQ不仅能测试像解码器这样的网络单元的效果,也能测量端到端的声音质量。它着重针对不同的信号退化原因,如编解码失真、错误、丢包、延时、抖动和过滤,给出-0.5~4.5的分值范围内的值。大多数情况下正常的分值范围在2~4.5之间。

PESQ总体思路是对源信号和通过测试系统的退化信号进行电平调整,使其到标准听觉电平,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的不同作为扰动(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算法将话音的频率、响度等物理特性与人类心理上的感知特性的对应关系用数学模型来表示,即用客观数学模型的评价来模拟主观的评价。该算法采用时频映射、频率弯折和响度弯折等方法,尽可能将语音中可以感知的特性在数学上完美的表达。在PESQ模型中,提取出的语音特征都是与主观感觉直接相关的。

PESQ算法主要缺点:处理CDMA编码(如EVRC)不够准确;在特定的GSM/WCDMA网络条件下过于敏感;不能处理超宽带语音信号。为了解决PESQ的问题,更好地适应4G/LTE时代的语音质量测试需求,ITU-T于2006年开始了POLQA的开发工作,2011年初正式发布为ITU-T P.863标准。主要特点可以覆盖最新的语音编码和网络传输技术,在用于3G、4G/LTE、VoIP网络时具有更高的准确性支持超宽带(50Hz~14kHz)高质量语音传输。

PESQ和POLQA算法的对比参见表3。

从主动式评价方法的发展历程以及各种方法的性

能来看,基于输入/输出的客观评价的一些方法与主观评价的相关度已达到了0.95左右,能够比较真实地衡量语音质量。

表3 PESQ和POLQA算法特性对比

	PESQ	POLQA	
Codecs	AMR EFR	AMR AMR-WB EFR EVRC EVRC-B EVRC-WB	iLBC AMB+ AAC Skype/SILK G.711 G.729
Reference Speech (sampling frequency)	8kHz	8kHz 48kHz	
Applications	POTS VoIP 3G	HD Voice Voice Enhancement Devices	Skype Calls Benchmarking CDMA and GSM

(3)被动式评价

被动式语音质量评价方法主要有两种:其一是直接从变化的IP网络损伤参数(如丢包、抖动和延迟)预测语音质量(如R-factor方法);其二是根据音频信号测量语音流量的语音质量(例如编解码器、回声、语言和/或讲话人),如ITU组织P.563标准(见图3)。被动式评价的目的是建立感知语音质量和网络或音频语音流相关参数的关系,从而通过网络参数或语音参数来评价感知语音质量。被动式评价并非真实测量最终用户的实际体验,而是通过抓取网络参数或语音参数通过数学模型“预测”感知语音质量。

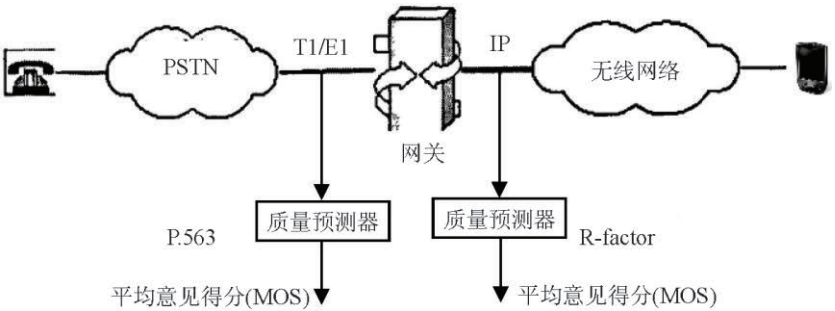


图3 被动式评价

3 Nomad HD 语音质量测试解决方案

思博伦通信研发的Spirent Nomad HD呼叫及语音测试系统能够针对任意网络上任意设备的高清语音和VoLTE服务,提供关键的语音质量分析能力。Nomad HD

可为高清语音质量测量和呼叫性能测试提供完整的测试包。该系统通过在高清语音和 VoLTE 语音质量测试中使用 POLQA 算法,可以在同一运营商的传统线路交换与分组交换语音之间进行切实可行的对比,并且可以对多种网络和技术进行跨越式对比。Nomad HD 支持带宽 20kHz 以上的高清语音和宽带 POLQA 评价算法。可以测试包括 OTT 语音在内的任何高清语音通话。云服务器部署在北美、欧洲、亚太等地区,包括美国、中国、中国台湾、日本、英国、德国等多个国家和地区,同时支持实验室测试和现网测试。

NomadHD 优势见图4,具体如下:

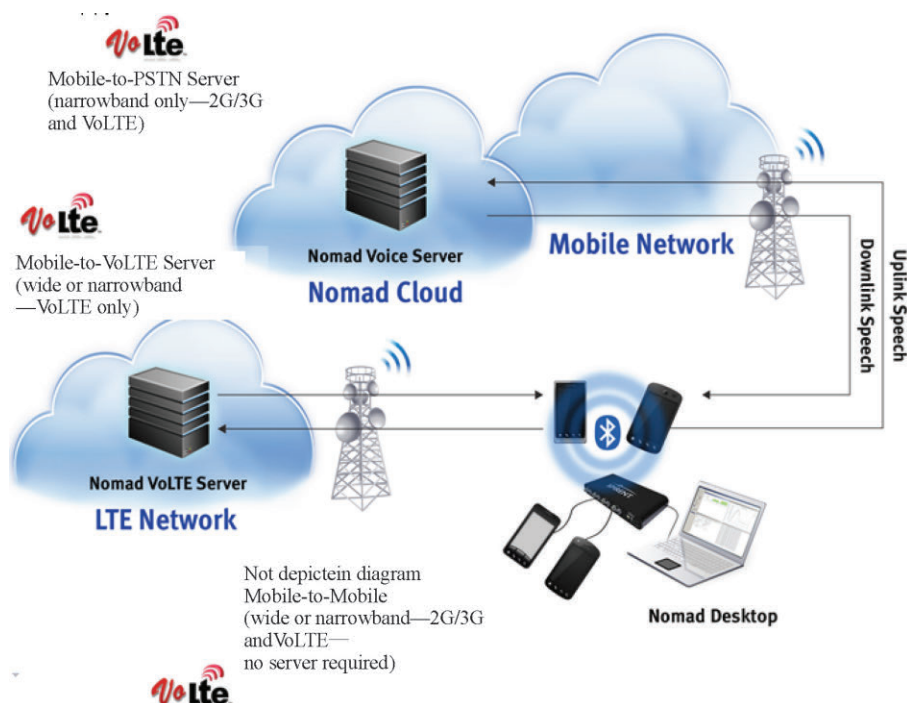


图4 Nomad HD 语音质量测试解决方案

●理解 VoLTE 服务部署需求,广泛应用于运营商准入测试和芯片、终端厂商研发测试中。

●客观比较 VoLTE 高清语音、OTT 语音,以及 2G、3G 语音服务质量。

●支持现网和实验室测试,支持所有制式、操作系统手机。

●同时关联分析 IP 网络对语音质量、语音延迟、呼叫性能的影响,帮助寻找系统瓶颈。

●作为 Fit4Launch 的预测试,帮助加快通过运营商准入测试的周期。

Nomad HD 架构见图5,具体如下:

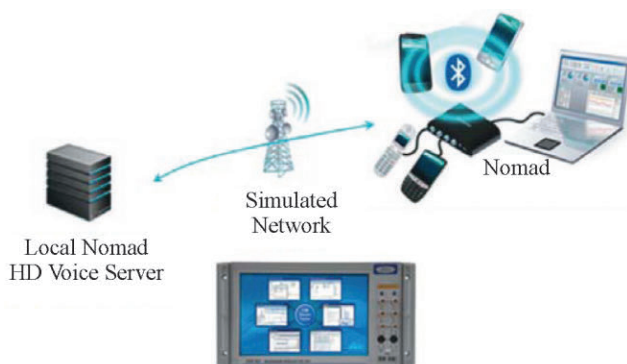


图5 Nomad HD 语音质量测试解决方案架构

●Nomad 桌面代理

Nomad 桌面代理包括便携 Nomad 硬件、PC 控制软件和 IP 分析模块。Nomad 硬件单元包括 4 路语音测试接口,可以通过耳机接头或蓝牙接口同时测试最多 4 部终端。PC 控制软件可以测量语音 MOS 分、语音延迟、起呼成功率、掉话率。Nomad 桌面代理支持终端到终端呼叫、终端到服务器呼叫等多种测试方式。Nomad 系统可以同时记录 IP 日志,结合语音质量进行 IP 分析。

●Nomad 语音服务器

Nomad 语音服务器由思博伦通信公司架设,通过 E1 或 T1 连接到 PSTN 网络。Nomad 语音服务器支持窄带语音测试,包括 VoLTE、2G/3G 或者 OTT 语音服务。

●Nomad VoLTE 服务器

Nomad VoLTE 服务器连接到运营商 IMS 核心网,通过 SIP 协议进行语音通话,支持 WB AMR (g.722.2) 语音编码方式。

Nomad HD 可以部署在实验室或者现网,可以与思博伦通信 CS8 网络仿真系统进行集成,在仿真网络环境下对语音质量进行测试。