

VoLTE高清语音质量自动化测试方案

罗德与施瓦茨（中国）科技有限公司

在LTE网络中，运营商最终将会使用基于IMS（IP multimedia subsystem）系统的IP语音技术为用户提供传统电话业务的替代服务。这一重大技术升级的优势在于网络使用高带宽的语音编解码技术承载“高清语音”（HD Voice，它的定义请参考GSMA官方网站上的说明），从而极大提高用户体验。人耳能识别的声音频率范围限制在20~20 000 Hz之间。传统固定电话网络和第二代数字移动通信系统（包括起源于欧洲的GSM和北美的IS-95），都只将频率范围在300~3 400 Hz之间的窄带声波信号作为信源编码的输入，其他频率成分则被滤除。这样做主要受限于两个因素，1）空中接口速率非常有限（如GSM的空口速率仅为270.833 kbps，且需要多用户时分复用）；2）无法实现高效率的语音压缩和编码。窄带语音的缺点非常明显，尽管通信双方的语意可以被识别，但是声音听起来单调、干涩、缺乏现场感。如信源编码时声音中的频率成分被保留的越多，那么用户听到的声音会越接近原声，感受更丰富、柔和、悦耳。在电话系统中，按照声音信号的频率范围，可将声码器划分为如表1所示的4种类型。

表1 电话系统中声码器的4种类型

类型名称	频率范围/Hz
窄带（narrow band）	300~3 400
宽带（wide band）	100~7 000
超宽带（super wide band）	50~14 000
全带宽（full band）	20~20 000

宽带与超宽带语音将主要应用于LTE网络，而全带宽语音作为超宽带语音的演进，将成为未来EVS（enhanced voice services）全高清语音业务的基础，覆盖声音全部频率成分的特性使得通过移动通信网络收听高保真音乐成为可能。

人对声音的感知是一个复杂的从物理现象到生理、心理作用的过程，即音源振动产生声音并以波的形式传播出去，声波进入人耳引起耳蜗内结构振动产生生物电信号，生物电信号传导到大脑刺激听觉神经使人感受到声音。人对声音的判别带有强烈的主观色彩。基于这一特点，声音相关的测试不仅要使用客观的检测手段（如元器件或整机的电压、频响等可定量参数的测量），还需要采用“听音”方式进行主观测试。传统的听音测试，由很多经过训练的“人耳”在标准环境下，聆听被测信号，按照一定的标准对声音质量的“好坏”进行打分完成，5分为“优秀”，1分则是“最差”。所有这些测试者打出的分数的平均，

就得到所谓“MOS”分数(mean opinion score)。理论上推算，VoLTE高清语音的最高编码速率等级，在静态传输条件下MOS分数可到达4分以上的效果，比传统2G/3G语音有明显的提高。

考察人对语音的主观感受，除了音质MOS分之外，还需考虑语音在整个通信过程当中延迟（Latency）的影响。喜欢看新闻节目的人，常会看到演播室和新闻现场远程联线的场景。连线距离如果很远，可看到演播室这边话已讲完，而连线那端的人长时间没有反应的现象，这就是延迟的影响，即由于通信距离过远，导致音视频信号需要长时间传输，从而使双方无法保持“实时”的交流。一般来说，通信系统中语音信号的延迟超过300 ms，双方就能够感受到通话的等待和暂停，引起不适。GSM、WCDMA、TD-SCDMA、CDMA2000这几种制式下，语音通过面向连接的电路交换网络传输，延迟比较小且固定。但是LTE网络是完全的分组包交换网络，IP包的突发特性和语音传输的要求本质上是矛盾的，加上IP协议的“发后不管，尽力而为”的特点，使VoLTE语音里延迟和通信质量之间的矛盾愈加突出。所以，从测试角度看，必须对VoLTE语音的延迟进行准

确测量，才有助于手机设计者提高VoLTE语音的整体性能。

上述人工听音方法现在已无法大规模应用，尤其在激烈的市场竞争下，厂家更无法投入巨大的人力进行此类测试。业界实际使用一种模拟主观听音的客观测试方法作为替代，这就是所谓PESQ[®]和POLQA[®]测试。PESQ是“perceptual evaluation of speech quality”的缩写，POLQA则是“perceptual objective listening quality assessment”的缩写。POLQA可以简单的理解为PESQ功能的升级和扩展，VoLTE语音质量测试使用POLQA方法，因此本文只介绍POLQA的测试（关于PESQ和POLQA测试的具体差别，建议读者可以参考阅读R&S公司编号为1GA62的应用文档当中有关内容）。这两种测试都是根据人耳工作原理，用物理和数学方法对人耳听觉功能建立PESQ或者POLQA分析模型。测试时将标准参考信号输入被测设备（以下简称DUT），获得经DUT后的有损信号输出，再将此输出信号输入PESQ/POLQA模型，与预存的参考信号进行对比，从而获得结果，如图1所示。POLQA模型除了能够获得MOS分数外，还可以通过比对有损输出和参考信号时间差得到整个通路的延迟，因此POLQA测量对VoLTE语音质量检测有着更加积极的意义。

真实的LTE网络环境复杂，在不同层面存在着多种影响通信质量的消极因素。而且LTE和2G/3G网络会长期

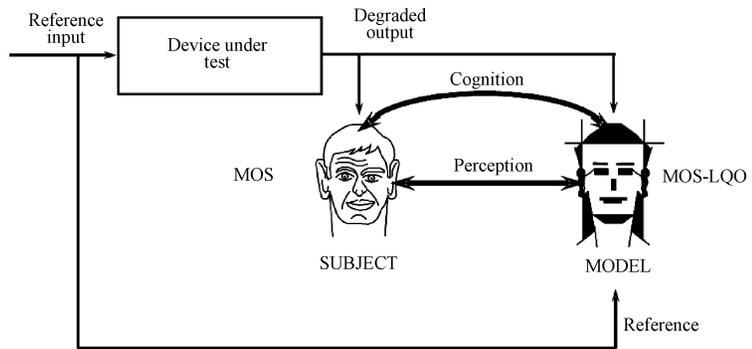


图1 POLQA测试原理示意

共存，用户在通话过程中难免会经历系统之间互操作的情况。此外，运营商会根据优化的需要约束DUT在网络中的一些行为。因此VoLTE语音质量测试不仅仅要覆盖静态传播场景，还要考虑到上述这些因素的影响。具体来说要着重进行如下测试：

1) 射频衰落场景下的MOS分数。在无线传播信道存在多普勒频移以及白噪声时，如果下行信号使用高阶调制方式和过大数据传输块尺寸，DUT将变得无法承受，进而引起MOS分数的严重恶化。此时网络通常需要改用较低的MSC (modulation scheme) 来提供更多冗余保护，从而保证音质。

2) IP层损伤场景的MOS分数和时延。因为VoLTE语音使用RTP协议传输，无后向纠错机制，IP层传输过程中的丢包会严重影响音质，使MOS分数下降，甚至导致连接中断。IP传输过程中很可能经过多个不同的网络设备，各设备间并没有时基同步机制，这使得连续的IP包之间在传输时间上出现波动，这一现象被称为“Jitter”（抖动）。为对抗抖动，DUT

在接收端都要设计一定大小的缓存，当连续到达的语言帧之间存在时间波动时，该缓存可存放若干连续语音帧，待语音稳定后再输出。较大尺寸的Jitter Buffer的会改善音质，即MOS分数提升，但会带来较大的处理延迟，引起通话者注意；反之较小容量的Jitter Buffer处理延迟小，但MOS分数优化有限。因此如何平衡延迟与MOS分数成为Jitter抑制的主要挑战（关于Jitter Buffer，请参考3GPP组织的规范26.448）。

3) 数据和语音同步传输时MOS分数。LTE网络下RTP语音和其他IP协议数据包同时传输是很常见的场景。DUT需要按照不同业务类型对数据包分别进行QoS处理。好的DUT设计应该是在MOS分数、延迟不恶化的要求下，最大程度保证其他数据业务的质量。

4) 切换场景下MOS分数和时延。切换包括有LTE网络内的跨小区切换、频点切换，还有通过eSRVCC方式完成的LTE->GSM，LTE->WCDMA等异系统间切换。切换动

作会造成DUT与网络的短时间内连接的丢失，并且带来额外的协议栈信令处理时间，所以切换会降低音质，增加延迟。

出于省电、扩大网络覆盖、降低信令开销等多种目的，运营商在LTE网络内可以使用cDRX（连接态下非连续接收），TTI Bundling（TTI绑定），SPS（半静态调度），RoHC（PDCP层包头压缩）等功能对DUT行为进行约束。开启这些功能对MOS音质以及延迟都有可能产生影响。

根据上面的说明可知，为保证质量，对手机的VoLTE语音质量需要进行全面的测试。因测试环境涉及到LTE基站与核心网，IMS网络与IP传输，音频编解码能力及标准音频源与分析等通信网络中的几乎所有重要网元与功能实体，而且测试中还需要随意设置基站功能，并精确的定量（比如延迟），所以整个测试非常复杂，无法在真实网络下开展，只能在实验室进行。

为了满足这类测试需求，罗德与施瓦茨公司开发了一套完整的VoLTE语音质量自动化测试系统，它具有功能全面，构成简单，运行速度快，人机交互友好等优点，十分适合开发和准入测试。该套系统仅由射频测试仪CMW500和音频分析仪UPV两台仪表，以及一台运行CMWrun自动化控制软件的电脑组成，如图2所示。



图2 R&S公司开发的完整VoLTE语音质量自动化测试系统

图2中上方的仪表是高精度音频信号分析仪UPV。它具有工作频率范围宽（直流10 Hz~250 KHz），测量范围广（0.1 μ V~110 V），灵敏度高（ ± 0.025 dB），噪声低等优点。它配备有双通道的独立音频源和分析仪，能同步分析DUT的上行和下行通路。在POLQA测试系统中，它的两路音频源负责播放按照3GPP规范要求制作的测量参考信号（或者运营商提供的测试语料文件），这种信号通常是一段混合了几秒钟男女声及静默的wav格式语音文件。按照测试方向的不同，参考信号分别发送至DUT的麦克风输入端（上行测量）和基站模拟器的音频编码器输入端（下行测量）；类似的，根据测试方向的差异，它的两路音频分析仪则分别接收来自网络音频解编码器输出端（上行测量）和手机的耳机输出端（下行测量）的有损声音信号，然后由内置POLQA测量软件测出该方向的MOS分数及通路的延迟。

图2中下方的仪表是无线通信综合测试仪CMW500。它是一款非常优秀的产品，自2007年10月面世以来，全球销量累计已达五万五千余台，已

经成为了行业内事实上的标杆，它的用户遍布有无线通信测试测量需求的各类机构。在POLQA测试系统中，CMW500起到了非常重要的作用，具体包括：

1) 模拟LTE/GSM/WCDMA/TD-SCDMA/CDMA2000等基站及核心网功能，允许手机注册，连接建立；根据需要执行LTE带内、带间、LTE到GSM或WCDMA等不同系统间的切换；LTE网络下可以提供cDRX，TTI Bundling，SPS，RoHC等多种实网复杂场景

2) 提供IMS服务和IP网络模拟功能，允许手机根据不同信令条件注册到IMS服务器，并建立VoLTE语音呼叫；提供IP层损伤模拟，包括可选择不同丢包率和Jitter时间；AMR声码器支持窄带、宽带到EVS等多种形式，最高支持23.85 kbps的编码速率；提供IP层抓包分析功能；内置不同类型IP服务器，如网页浏览、FTP、Video Streaming等，可模拟VoLTE语音和数据业务并发场景

3) 上下行双向音频硬件接口，提供电压和阻抗匹配，并可实时更改编码器速率，可真实还原实网下DUT根据网络要求更改语音速率的场景

4) 可准确测量网络侧处理造成的延迟。此功能不仅对VoLTE有效，对GSM，WCDMA，CDMA2000，TD-SCDMA等电路交换业务，CMW都可以计算出每一个语音帧（20 ms）在网

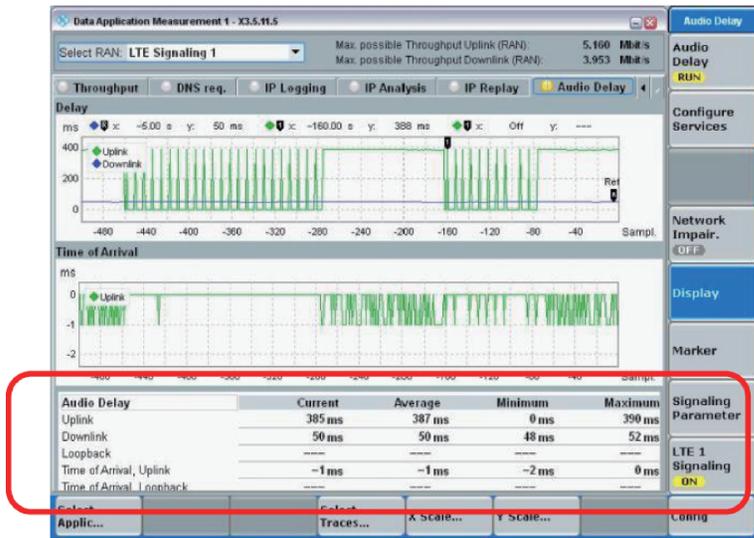


图3 VoLTE下的网络侧延迟结果示意

UPV_SpeechQualityMeasurement: POLQA Measurement

Start Measurement: POLQA - Downlink, Super-Wide, Single-Mode
1/1/5
← MOS=3.9342
2/1/5
← MOS=3.9143
3/1/5
← MOS=3.9703
4/1/5
← MOS=3.8698
5/1/5
← MOS=3.9111

Test Item	Lower Limit	Avg. Delay	Measured	Status
Measurement (Ref. File: 'C:\UPV\Config\ref\PolqaRef48000.wav' - Downlink, Super-Wide, Single-Mode)				
POLQAMeasurement 001	---	178.2 ms	3.9342	---
POLQAMeasurement 002	---	178.1 ms	3.9143	---
POLQAMeasurement 003	---	177.5 ms	3.9703	---
POLQAMeasurement 004	---	177.8 ms	3.8698	---
POLQAMeasurement 005	---	177.8 ms	3.9111	---

Test Item	Min MOS	Max MOS	Avg MOS	StdDev
Statistics (Ref. File: 'C:\UPV\Config\ref\PolqaRef48000.wav' - Downlink, Super-Wide, Single-Mode)				
POLQA Measurement 5 bps/(g)	3.8698	3.9703	3.9199	0.0328
Median	---	---	3.9143	---
03rd Percentile	---	---	3.6747	---
10th Percentile	---	---	3.6803	---
20th Percentile	---	---	3.9111	---
30th Percentile (Median)	---	---	3.9143	---
75th Percentile	---	---	3.9342	---
90th Percentile	---	---	3.9559	---
97th Percentile	---	---	3.9650	---

Test Item	MOS Limit	Percentile Limit	Calculated	Status
Evaluation (Ref. File: 'C:\UPV\Config\ref\PolqaRef48000.wav' - Downlink, Super Wide, Single Mode)				
POLQA Measurement 80.00th Percentile > 3.00	3.00	80.00 %	100.00 %	Passed

图4 CMWrun提供详细测量报告示意

络侧的延迟，而不是给出一个模糊的经验值，VoLTE下的网络侧延迟结果如图3所示。

系统中的电脑将会安装CMWrun自动化测量软件，它通过运行预先编制好的测试脚本，控制CMW500和UPV，进行各种测试场景模拟及POLQA测量。CMWrun提供的测试脚本非常灵活，可根据需要进行修改，

以满足个性化的测试要求。CMWrun还可通过安卓系统ABD接口控制DUT，执行开关飞行模式，接听电话等动作，做到完全的无需人为干预的自动化测试，极大的解放人力。因VoLTE测试需验证的场景很多，不少测试用例（如切换类测试）对参考信号的播放时间与测量时间窗口之间的配合有严格要求，所以手工测试效率

低下且不准确，而CMWrun自动化测试的方法可以精确的控制测试过程及定时，因此非常适合这类测试。测试结束后，CMWrun还可以提供多种格式的详细测量报告。

除了上述优点之外，R&S测试方案中延迟的测量也有独到之处。首先系统使用CSS（composite source signal）信号作为延迟测量的参考信号。该信号由语音、噪声和静默组成。信号中语音的作用是让DUT接收端的Jitter Buffer能够尽快收敛以便执行延迟测试。而时间的定位则是根据伪随机噪声信号良好的互相关特性来完成，当有损信号与参考信号做互相关运算得到尖峰时，即获得准确的时间信息。静默的添加则使得整个测试信号更加接近真实通话场景。CSS信号在3GPP规范26.132以及ITU技术要求P.501当中都有详细规定，具有权威性和准确性。如图5所示为R&S方案中使用的CSS信号的时域构成。

另一个独到之处就是R&S公司的测试系统提供单端和环路两种延迟测量的方法。真实通信中，延迟来自于网络和手机各自的处理时间，以及空中接口的传输时间。实验室测试时，空口传输时间可忽略不计。UPV的POLQA模型计算得到的是整体延迟，该时间减去仪表侧精确测量的网络内部延迟，即可获得DUT侧的延迟。运营商往往关心整个环路中的延迟性能，它包括了上行和下行两个方向；然而从手机厂商角度看，DUT接收侧

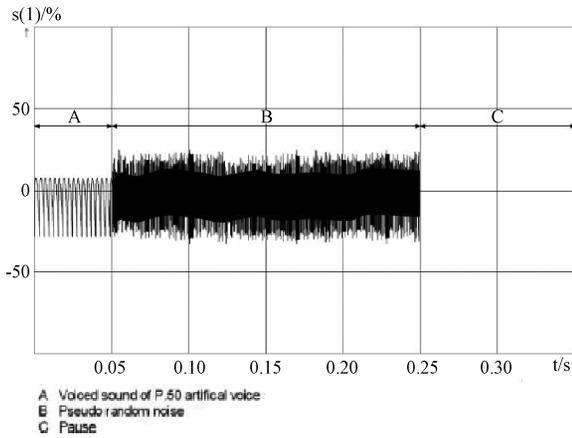
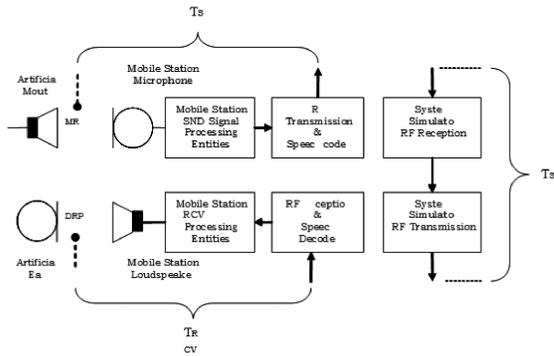
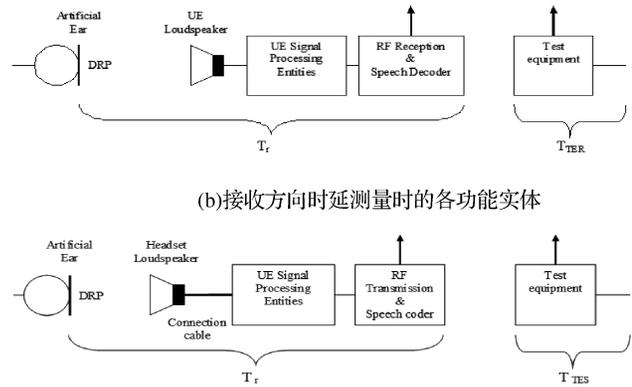


图5 R&S方案中使用的CSS信号的时域构成

延迟更加具有检测的意义。这是因为DUT里延迟主要来自于接收通路里的Jitter Buffer, 缓存容量的大小决定DUT侧延迟的表现。正是基于用户不同的关切, R&S公司的POLQA测试系统提供了单端和环路两种延迟的结果, 向客户提供了更为深入的分析数据。上面两种延迟测量方法, 都是有严格依据的, 在3GPP规范26.132中有详细的描述, 如图6所示。



(a)发射和接收方向联合时延测量时的各功能实体



(b)接收方向时延测量时的各功能实体
(c)接收方向时延测量, 被测手机连接外置扬声器时的各功能实体

图6 单端和环路延迟测量方法示意

功能上看, R&S公司的测试系统确实是VoLTE语音质量测量的一大利器。现实中, 该系统也已经在为许多不同类型客户提供服务了, 尤其是R&S公司和多家大型移动运营商, 如中国移动、中国联通、美国AT&T、美国Verizon等对不同芯片, 不同品牌

的DUT进行了大量的联合调试, 效果显著。根据运营商的需求, R&S公司提供了成套的定制CMWrun测试脚本用于准入测试, 有效推动了整个产业界的技术升级和发展。而且, 对手机厂商来说, 一个非常现实的问题就是如何降低成本。R&S测试系统有一个

突出的优势, 就是CMW500和UPV的保有量非常多, 而且CMW的扩展性能非常好。客户升级时无需花费太多的时间和金钱, 甚至只需购买CMWrun软件选件, 即可快速开展相关测试, 这为保护客户既往投资起到了十分积极的作用。