Validating voice over LTE end-to-end

As LTE becomes more widely adopted, interest in how to carry voice over LTE is growing rapidly.

MICHAEL ANEHILL, MAGNUS LARSSON, GÖRAN STRÖMBERG, ERIC PARSONS

Unlike previous 3GPP wireless technologies, LTE has no circuit-switched bearer to support voice, so carrying voice over LTE requires a migration to a voice over IP (VoIP) solution – commonly under the umbrella of IMS. Until this migration occurs, LTE-capable handsets need to revert to 2G or 3G for voice calls: an approach that is not ideal in the long term. Driven by major operators, the voice over LTE (VoLTE) ecosystem is maturing rapidly and tracking well to target commercial timelines.

Several VoIP solutions for wireline IP networks have existed for some time now, whereas solutions for wireless networks are less mature. To facilitate the necessary migration to VoIP, extensive work needs to be carried out so that the mobile-broadband capabilities of LTE networks reach the same coverage and reliability levels, as is possible for circuit-switched voice in 2G and 3G networks. Similarly, the implementation of LTE devices and the way they interact with the network must mature so that performance levels are equivalent to those achievable with 3G handsets.

When compared to its predecessors, LTE differs in many ways. Most significantly, the nature of dynamic scheduling coupled with variable retransmissions, interspersed with inter-cell handover, can introduce significant jitter. To cope with this, devices require sophisticated jitter-buffer solutions. LTE introduces new hardware and software interfacing between speech sampling, RTP/IP packet generation, header compression and modem scheduling functions.

For VoLTE solutions to reach the level of maturity required for commercial deployment, a number of lab, field and market trials must be conducted. Lab trials have been under way at Ericsson for nearly two years, and over-the-air (OTA) field trials for about a year. The most recent field trials were carried out using commercial-track form-factor-accurate smartphones, as these devices are approaching the levels of integration required for industry players to fully understand how target solutions will perform end-to-end (E2E). Some of the trials included video-telephony as an enhanced communication service.

BOX B

MOS-LSQM

Mean opinion score – listening quality subjective mixed (MOS-LQSM) is the unit for subjective grading in a MOS test that includes both narrowband and wideband voice.

BOX A Terms and abbreviations

2G	2nd-generation wireless telephone technology
3G	3rd-generation wireless telephone technology
3GPP	3rd Generation Partnership Project
AILG	air interface load generator
AMR-NB	Adaptive Multi-Rate Narrowband
AMR-WB	Adaptive Multi-Rate Wideband
AVC	advanced video coding
CBP	Constrained Baseline Profile
CBR	constant bit-rate
CS	circuit-switched
CSCF	Call Session Control Function
DL	downlink
E2E	end-to-end
ECM	EPS Connection Management
EPC	Evolved Packet Core
EPS	Evolved Packet System
FER	frame error rate
fps	frames per second
FTP	File Transfer Protocol
G.114	ITU-T one-way transmission time aspect of transmission quality standard
H.263	ITU-T video compression standard – predecessor to H.264
H.264	ITU-T video compression standard equivalent to MPEG-4 AVC (Part 10)
HARQ	hybrid automatic repeat request
HSS	Home Subscriber Server
IMS	IP Multimedia Subsystem
IP	Internet Protocol
ISC	IMS service control/IMS service control interface
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
JBM	jitter-buffer management
LTE	Long Term Evolution
MME	Mobility Management Entity
MOS	mean opinion score
MOS-LSQM	mean opinion score – listening quality subjective mixed
MPEG	Moving Picture Experts Group
MTAS	Multimedia Telephony Application Server
OTA	over-the-air
OTT	over-the-top
PDN-GW	packet data network gateway
PLR	packet loss ratio
PS	packet-switched
QCI	QoS class identifier
QoS	quality of service
QVGA	Quarter Video Graphics Array
RAN	radio-access network
RF	radio frequency
RTP	Real-time Transport Protocol
SGW	service gateway
SINR	signal-to-interference-plus-noise ratio
SIP	Session Initiation Protocol
SPD	speech path delay
TCP	Transmission Control Protocol
UE	user equipment
UL	uplink
VoIP	voice over IP
VoLTE	voice over LTE
WCDMA	Wideband Code Division Multiple Access

This article covers the key performance characteristics of voice and video-telephony solutions according to the GSMA IR.92 and IR.94 specifications, from a control- and media-plane perspective. The field-trial results show that the ecosystem is maturing quickly, with good E2E performance observed. Certain devices exhibited implementation weaknesses that were subsequently corrected. Although more rigorous testing needs to be carried out, current results suggest that the ecosystem is on track to meet the deployment timeframe desired by leading VoLTE operators.

Characteristics: media quality

User perception of the overall merits of a VoLTE call is determined by the time spent setting it up, and the quality of the speech (and eventual video) aspects of the session.

Speech quality varies according to a number of factors relating to the terminal and the network. For terminals, speech quality depends on the performance of the microphone and speaker, as well as the functionality to handle acoustic echo, background noise, compensation of speech level, and especially the speech codec and jitter-buffer management (JBM).

Video quality is also determined by a combination of factors, even more of which are terminal-related than with speech. The quality of JBM, the camera and display, the appropriate video codec, the settings of frame-rate and image fidelity for the format used, as well as the adaption of speech versus video delay (lip-synch alignment) all affect user-perceived video quality.

In the absence of media gateway functions, the network impacts speech and video quality in terms of transport and mobility. The network mainly affects quality through available bandwidth, packet loss (which causes speech/video frame error), and delay as well as interruption time at handover.

Except for the speech- and video-path delay factor, the impact of all other speech and video quality factors is the same, regardless of whether the call is handled over a circuit-switched (3G/2G) or a packet switched (LTE) network. For calls over CS, the delay is fixed on both an E2E and on a node level for their entire duration. For calls over PS, a large amount of jitter is introduced in the LTE radio network – as a result of speech- and video-quality and network-capacity/coverage optimization. Despite the introduction of jitter, the target is still to provide a fixed E2E delay for PS-call users.

The two factors directly related to the LTE radio network that impact the quality of speech and video are packet delay and packet loss rate. These factors are, in turn, highly dependent on the capability of the LTE radio network to perform link adaptation and packet scheduling under various load and interference conditions, while maintaining efficient use of the radio-interface resources. (link adaptation refers to the the ability to change transmission mode based on the radio condition of the terminal and packet scheduling refers to the prioritization of packets between different terminals and IP flows). The use of L2 HARQ retransmission in LTE can create significant additional jitter in packet delivery; some packets may be delivered on first transmission, while others may use up to the maximum configured number of retransmissions to be delivered successfully. The dynamic nature of the packet scheduler may also introduce significant jitter, at times transmitting packets as soon as they arrive and at other times, very close to the maximum delay budget (taking into account the HARQ retransmissions). In some cases, the handling techniques lead to the late arrival of packets, which – from a quality perspective – results in a lost (faulty) speech or video frame.

The JBM in the terminal needs to be designed so that the user perceives delay variations caused in the radio network as negligible.

Call setup time

The elapsed time between the moment the originating UE sends an SIP INVITE to initiate the call and when 180 RINGING is received by the terminating UE is defined as the call setup time.

Targets – media quality

The overall quality target for VoLTE is for voice calls to be perceived as on a par with, or better than, a 3G call. Using WCDMA as the 3G benchmark, the default target is for speech quality to be greater than 3.5 MOS-LQSM on a call-by-call basis, as illustrated in **Figure 1**.

FIGURE 1 **Speech quality versus frame error rate**

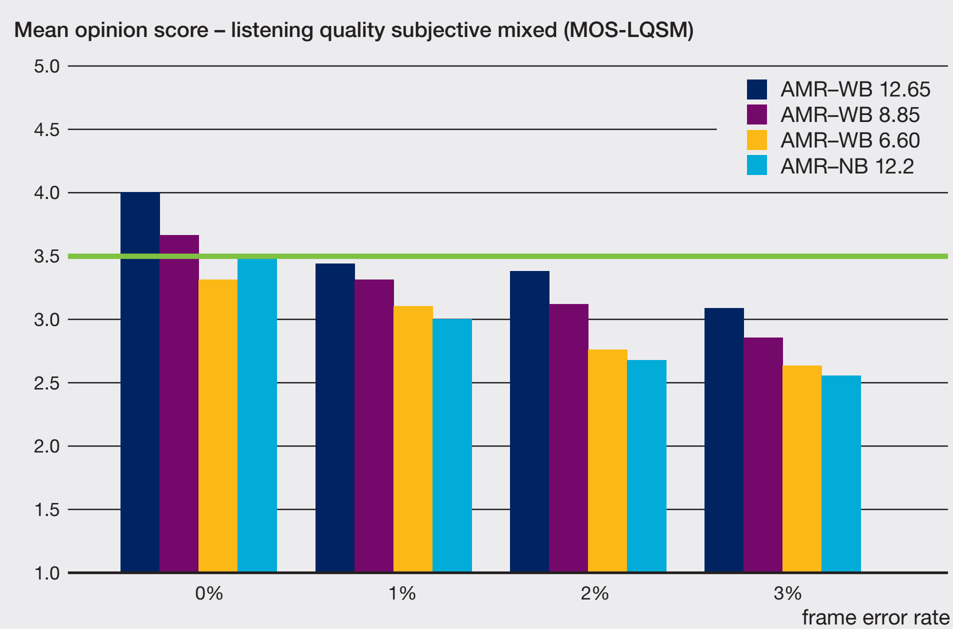
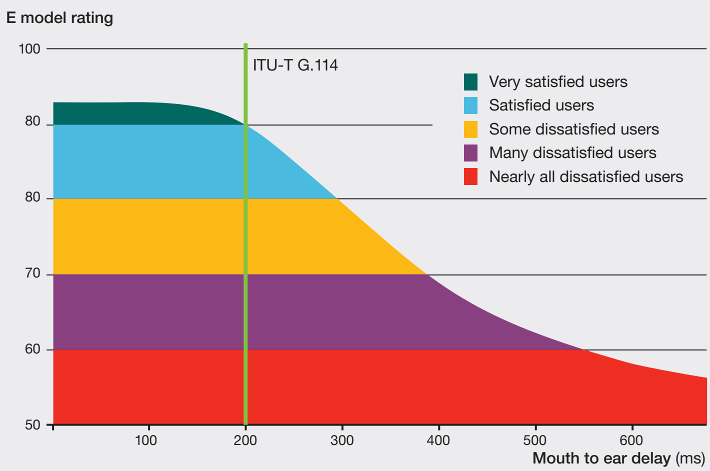


FIGURE 2 Speech quality versus mouth-to-ear delay



The focus here is the network, so in terms of transport parameters, this means a frame error rate (FER) of less than 1 percent and a mouth-to-ear delay of less than 200ms (reflecting a UE to UE call), as shown in **Figure 2**.

To achieve the same speech quality for a VoLTE call as for a WCDMA call, the target performance levels are:

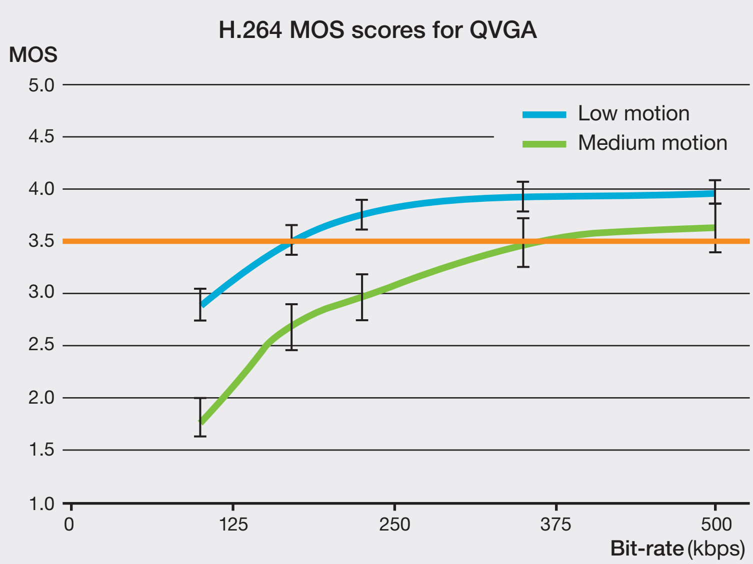
- **a packet loss ratio (PLR)† of less than one percent; and**
- **a mouth-to-ear delay of less than 200ms – of which nearly half of the time is dedicated to handling jitter from the LTE radio network‡.**

For a VoLTE video call, the overall target is to deliver video quality that is much better than that provided by WCDMA. Assuming the H.264 Constrained Baseline Profile (CBP) codec is used, the performance targets – as illustrated in **Figure 3** – are:

- **a video bit-rate of more than 350kbps§;**
- **a PLR* of less than 0.1 percent – corresponding to a visible error no more than once every 20 seconds; and**
- **a camera-to-display delay of less than 400ms – of which more than half is dedicated to handling the jitter introduced by the LTE radio network††.**

Basically all packet loss and introduced jitter come from the LTE air interface.

FIGURE 3 Video quality versus video bit-rate for low and medium motion content



BOX C

† For voice, the recommendation is for a 1:1 relationship to exist between speech frames and RTP packets – in other words, PLR should be the same as FER.

‡ In reality, 99 percent of the RTP packets will have been delivered to the UE within the budgeted time.

§ Note that only video bit-rate is mentioned, implying that transport bit-rate must be a little higher.

* For video, multiple RTP packets – at least two at 350kbps – are needed to transport one video frame, implying that the relationship between lost frames and packets for video is not 1:1.

†† In reality, 99.9 percent of the RTP packets will have been delivered to the UE within the budgeted time.

Call setup time

The target for VoLTE-to-VoLTE calls is to have shorter setup times than WCDMA-to-WCDMA calls. The target setup time for VoLTE calls is to be below three seconds – the setup time for a WCDMA call is typically more than this. If the terminating UE is in an ongoing data session, the setup time for the VoLTE call will be significantly lower as no paging is needed.

Measured VoLTE voice and video performance

During the second half of 2011, Ericsson conducted a VoLTE field trial, which was further complemented by LTE OTA‡‡ lab tests. The objective for both types of tests was to characterize and assess the quality and performance for voice and video calls over LTE in varying radio, load and mobility conditions.

The tests used the latest developments for facilitating VoLTE, including the Ericsson LTE solution, a multi-vendor IMS solution and pre-commercial UEs with VoLTE capability. All were in a commercial deployment configuration.

Test conditions

To ensure that the tests reflected real-world conditions, each case was executed with varying RF and mobility scenarios. The following conditions were used in the trial:

- **good: SINR of more than 16dB;**
- **medium: SINR of 2dB to 16dB;**
- **poor: SINR of less than 2dB;**
- **single-sector mobility: the drive route for this test condition covered varying locations within the cell/sector, from near the eNodeB site to the cell edge;**
- **intra-eNodeB handover; and**
- **inter-eNodeB handover.**

The originating UE was moving about in locations with various RF conditions while the terminating UE was stationary in good RF conditions.

Background traffic load

To enhance the approximation to real-world conditions, all test cases were executed with background load generated in the same cell as the test calls. Three types of background load were simulated: TCP traffic, load on the calling UE, and eNodeB load.

The iPerf tool executing on four devices (PCs with dongles) generated the TCP traffic. All devices were stationary, positioned in locations with varying RF conditions, generating DL traffic with two devices generating UL traffic. Typical throughput measured 4-5Mbps on the DL and 0.5-1Mbps on the UL. A Linux server connected via the SGi interface in the core network site served as both server and client for each UE.

An FTP client was used to generate load on the calling UE. This client downloaded a large file from the Linux server in the core network site while the test call was in progress. Throughput varied depending on RF conditions, and was typically in the range of 1-5Mbps.

The AILG tool was used to generate load on the eNodeB. The target load level was set to 100 percent, to ensure that the eNodeB was under maximum load regardless of the actual traffic level.

QoS configuration

All test cases were executed using QCI 5 for SIP signaling, QCI 1 for voice and QCI 8 for default/background traffic. For the purposes of this study, video was carried on a dedicated non-guaranteed bit-rate bearer.

User equipment

Two types of UE were used in the trial. A leading terminal vendor provided a pre-commercial handset with VoLTE capabilities that was used as the originating and terminating device in all test cases. The same vendor provided the LTE dongles for the PC laptops generating background data traffic.

Codecs

The handset provided by the terminal vendor offered two codec selections for voice: AMR-NB at 12.2kbps and AMR-WB at 12.65kbps. To benefit from the full capability and voice quality of mobile broadband, only AMR-WB at 12.65kbps was used throughout the trial.

For video telephony calls, two choices of codec were available: H.263 and the more recent H.264/AVC (MPEG-4 Part 10). Again, to benefit from the full capability of the network, the H.264 codec was chosen as a video codec throughout the trial.

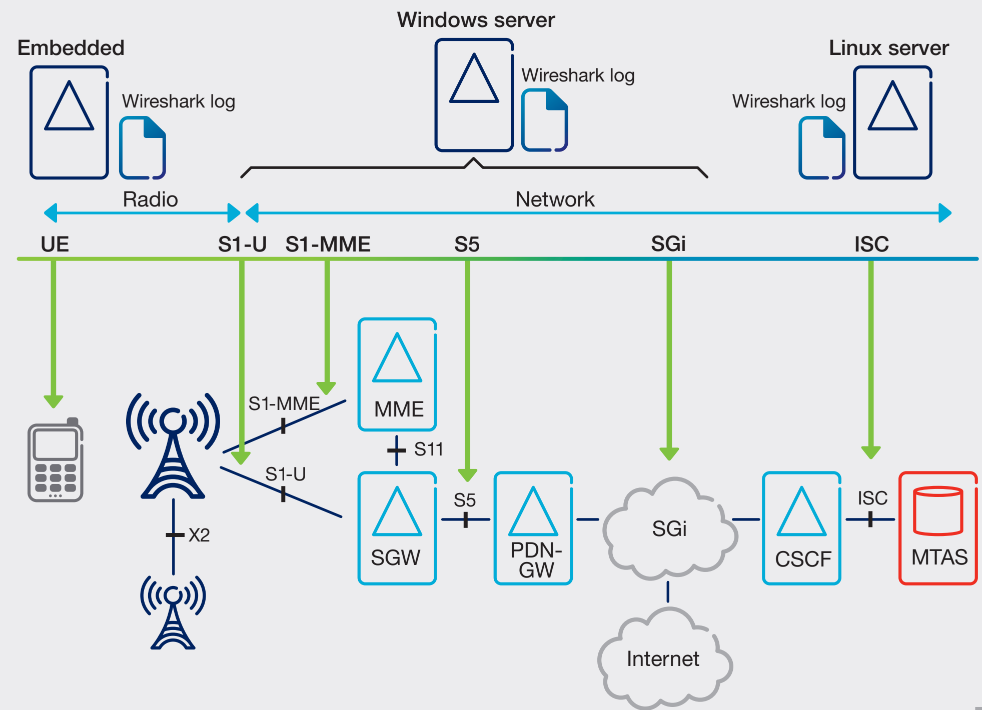
The H.264 codec used in the trial offered a limited set of parameter configurations. The following settings were used for all video test cases:

- profile: Constrained Baseline Profile;
- compression level: 1.2 (384kbps max video bit-rate);
- resolution: QVGA 320x240;
- frame rate: 20fps; and
- bit-rate control: constant bit-rate (CBR).

Data collection and analysis tools

A variety of tools for data collection were used during the trial, the choice of which was determined by the defined KPIs and by the test-case design for measuring them. A combination of »

FIGURE 4 Data capture points



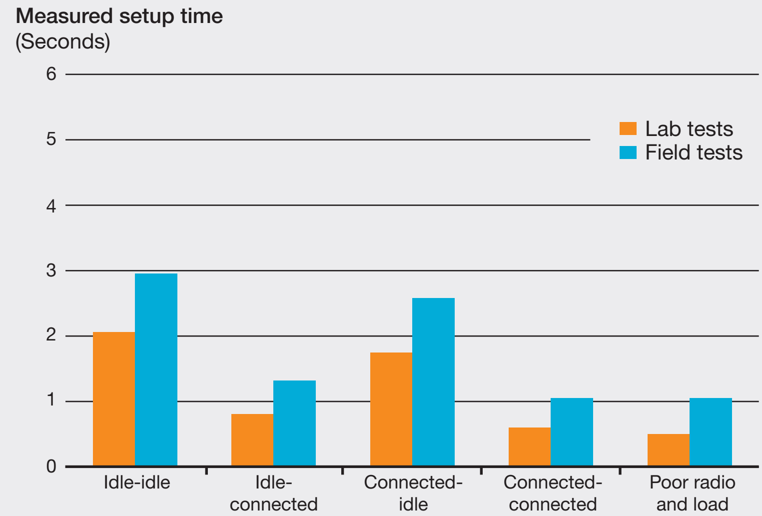
BOX D

‡‡ Over-the-air test system, indicating that the UE is connected to the lab network via the air interface, not via a coaxial cable.

Table 1: Test cases executed in the lab and field trials

VoLTE session setup over QCI 5 Good RF	VoLTE session setup over QCI 5 Poor RF with background load
UE-A ECM-IDLE UE-B ECM-IDLE	UE-A ECM-CONNECTED UE-B ECM-CONNECTED
UE-A ECM-IDLE UE-B ECM-CONNECTED	
UE-A ECM-CONNECTED UE-B ECM-IDLE	
UE-A ECM-CONNECTED UE-B ECM-CONNECTED	

FIGURE 5 **Measured call setup time**



BOX E

§§ Since PLR and FER here is the same for voice, the term PLR is used for the results.

** The UE used in the field trial was a second-generation handset. In the lab trial, a laptop with a non-optimized client combined with a dongle was used.

commercial and internal tools was used as a result:

- Wireshark, an IP packet analyzer;
- Audacity, an audio analysis tool for recording and analyzing analog audio samples;
- terminal RF logs; and
- internal Perl scripts to calculate skew/clock drift, jitter, RTP latency and PLR.

Network data capture points

A Windows 2003 server located in the core network site captured Wireshark traces on the S1, S5 and SGi interfaces and a Linux server captured Wireshark traces on the ISC interface. These traces were used primarily for call-setup latency measurements. For most of the test cases, the traces captured from the embedded Wireshark in the handset UEs were used for analysis. **Figure 4** illustrates the capture points for each interface.

Test results

Call setup time

For VoLTE, call setup time is measured from the moment the originating UE sends the INVITE to the moment RINGING is received from the terminating UE.

The call-setup-time results from the lab and field trials meet the targets for VoLTE, and are significantly shorter than WCDMA setup time for non-idle-idle cases. The systematic difference between lab and trial can be attributed to the UE and the IMS deployment, as these two components contributed to a greater part of the call setup time in the field trial than in the lab trial.

Speech quality

If the test results show a PLR§§ of less than one percent and a mouth-to-ear delay of less than 200ms, the quality target for speech has been met. Consequently, the results of these parameters are used in this article to determine whether quality targets were met.

Both mouth-to-ear delay and PLR are measured over the QCI 1 bearer, which has been optimized for VoIP.

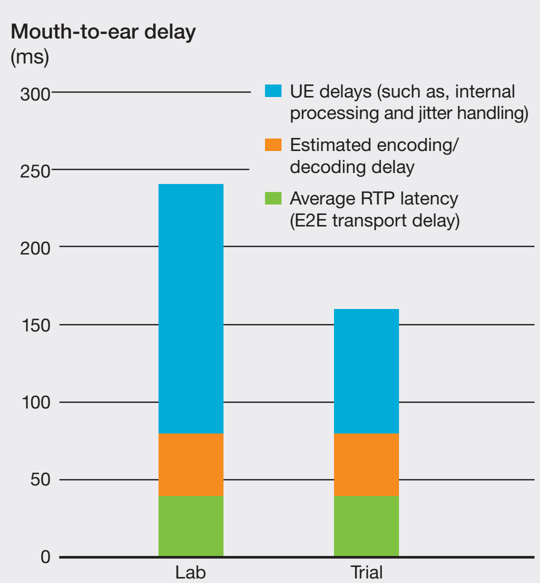
As illustrated in **Figure 5,** the target mouth-to-ear delay – less than 200ms – was met in the field trial but not in the lab trial. The main reason for a higher mouth-to-ear in the lab trial was the UE**. The UE factors contributing to this long delay were processing from sound-card implementation and prioritization of other functions.

The mouth-to-ear delay results, combined with a very low PLR shown in **Table 2** – measured to be about 0.4 percent at the cell border – indicate that VoLTE speech quality meets its targets (better than 3.5MOS–LQSM), and will offer very good performance in the entire cell.

Inter-arrival jitter

When comparing voice over packet-switched options with voice over circuit-switched options, the main issue is transport – where the packet-switched radio network introduces a significant amount of jitter. The inter-arrival jitter

FIGURE 6 **Measured mouth-to-ear delay**



measured in the lab, for voice calls and video telephony calls in the field trial with varying RF conditions, is calculated according to the formula in **Box F**.

The values for inter-arrival jitter shown in figures 7, and 8 indicate the delay – the difference between actual arrival time and expected arrival time – of the current package with respect to the previous package. A value of 0ms indicates that both packets have the same delay; a positive value indicates that the delay for the current packet is longer. A negative value indicates that the delay for the current packet is shorter than the preceding one.

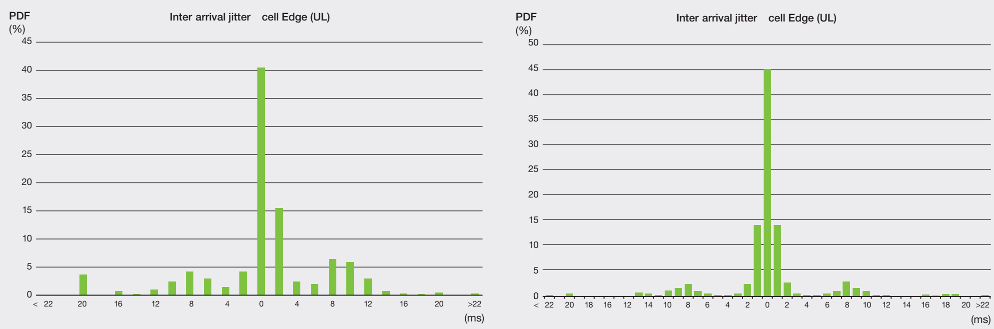
The inter-arrival jitter for a voice call at the cell border is illustrated in **Figure 7**. The inter-arrival jitter measured in the trials is quite adequate, as it is well within the delay budget specified as the target for speech quality. These results show that LTE radio-network performance is more than adequate for providing good VoLTE speech quality.

Video quality

The inter-arrival jitter for a video telephony call is illustrated in **Figure 8**. Naturally, the measured inter-arrival jitter is higher for video telephony than for a voice-only call, due to the larger packets sizes and the associated scheduling variability. However, since the time allocated for handling jitter is much greater for video telephony than for voice-only telephony, the inter-arrival jitter measured should result in a PLR of less than 0.1 percent – as targeted.

The video was run at a CBR of 384kbps with 20fps, which is in line with the recommended target. This, combined with the low PLR indicated from the inter-arrival jitter results, indicates that a high-quality video telephony service can be offered over LTE.

FIGURE 7 **Voice UL inter-arrival jitter measured in the lab trial (left) and in the field trial (right)**



BOX F

Inter-arrival jitter

The inter-arrival jitter presented was calculated for two consecutive RTP packets: i-1 and i, as:

```
D(i, i-1) =
arrival_time(i) –
arrival_time(i-1) –
(RTP_timestamp(i) –
RTP_timestamp(i-1))/
divider
```

Where the divider is 16 for AMR-WB and 90 for H.264

Conclusion

In early launches of VoLTE, either in the form of market trials or commercial availability, the need to ensure a good user experience is paramount. Given the trend toward data-rich applications, voice now represents a smaller fraction of total device usage. It does, however, remain an essential capability, and user expectations for voice are very high. If voice services do not deliver the necessary level of quality and reliability, users will revert back to existing circuit-switched options or, in some cases, simply rely on over-the-top solutions.

Clearly, the performance of the UE will play an even greater role for services provided over packet-switched networks than over circuit-switched. To meet VoLTE performance targets, JBM for the UE needs to meet specifications, and the delay contribution from other (post-decoder) UE internal processing outside the normal functionality must be negligible.

Efforts to validate VoLTE from an E2E perspective, which have been under way for more than a year, have the key objective of retaining users. LTE was launched commercially in late 2009, and the first VoLTE functional tests were conducted at that time. In early 2010, substantial lab trials were undertaken, first to begin understanding how to measure voice quality and then to validate it in the first commercial systems. In Germany in early 2011, the first solution for wireless voice over IP using an LTE network was commercially deployed, validating many of the QoS and scheduler enablers that are required for VoLTE. It is a testament to the sound design of LTE that this deployment could occur so quickly after the introduction of a new mobile-broadband wireless technology. Finally, in mid-2011, the first E2E OTA field trials of VoLTE were possible using pre-commercial form-factor-accurate devices – and these trials form the basis of this article.

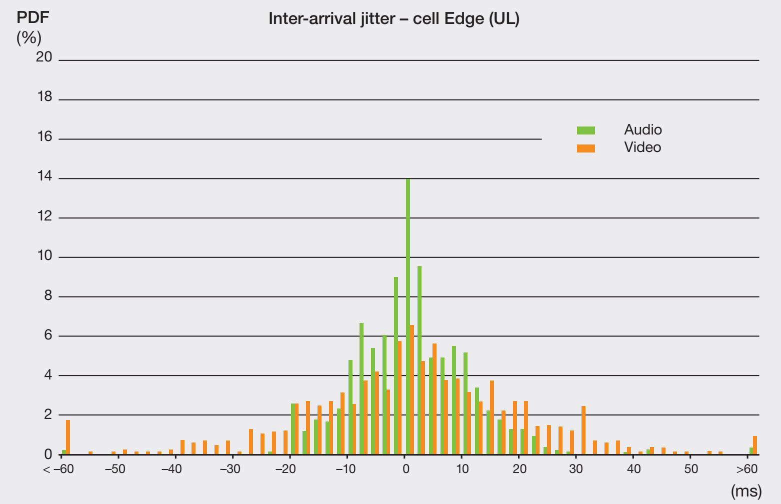
The industry is progressing rapidly in terms of KPIs, measurement practices and, ultimately, the performance of early solutions. While some improvements remain to be made, particularly in relation to UE JBM and implementation delays from other UE internal processing, the current performance level is already close to, and in some cases superior to, those specified in the target objectives. In parallel, it is likely that LTE schedulers will become more sophisticated in terms of managing mixed voice and data traffic, while ensuring good battery performance.

In short, VoLTE works and is ready for widespread adoption. It is well defined in 3GPP and GSMA standards, it has been implemented and has now been verified. Due to its telecom-grade implementation, which supports stable quality levels throughout the entire network regardless of load, VoLTE should, under normal network conditions, outperform unpredictable OTT voice and video services.

Table 2: Measured PLR

	Lab	Field
PLR	<0.4%	<0.4%

FIGURE 8 UL inter-arrival jitter measured in the field trial – video telephony



Michael Anehill

joined Ericsson in 1992. He has worked with speech quality since 1996, starting with TDMA followed by PDC, GSM and WCDMA, and is now focusing on multimedia telephony. He holds an M.Sc. in applied physics and electrical engineering from Linköping University, Sweden.

Magnus Larsson

started with Ellemtel (now wholly owned by Ericsson) in 1991 to work on the development of the next generation AXE system. He has spent more than 25 years in the telecoms industry. He is currently a senior specialist in the area of IMS characteristics at Product Development Unit, Core and IMS in Stockholm, Sweden.

Göran Strömberg

is a system manager at Mobile Broadband Engagement Practices in Region North America. He has recently been working with VoLTE performance and characteristics in networks with commercial deployment configuration. He holds a B.Sc. in applied mathematics and computer science from Uppsala University, Sweden.

Eric Parsons

is a strategic product manager for telephony over LTE. He has held a number of senior technical leadership positions in wireless product architecture and standards focusing on end-to-end solutions with more than 20 years industry experience. He has published over 20 refereed publications and numerous patents. He has a Ph.D. in Computer Science from the University of Toronto, Canada.