Voice/Video Quality Measurement for LTE Services

Yingrong Coral Sung, Yung-Sheng Ho, Yi-Bing Lin, Jyh-Cheng Chen, and Herman Chung-Hwa Rao

Abstract

LTE is a broadband wireless communication standard with the core network technology including voice service, which will be enhanced and evolved into 5G. In this article, we elaborate on performance issues of VoLTE and ViLTE. We investigate VoLTE voice quality with measurement experiments performed in the ITRI VoLTE testing environment. We also present the multimode and multiband interworking ViLTE call with a trial in a FET commercial mobile network. Our study provides references for the operators to measure the quality of VoLTE and ViLTE services.

Introduction

Long-term evolution (LTE) is a broadband wireless communication standard that utilizes IP Multimedia Subsystem (IMS) to provide voice and multimedia services [1]. In future 5G networks, LTE is considered one of the radio technologies, and Voice over LTE (VoLTE) service will still play an important role in 5G voice services [2].

Figure 1 illustrates a Nokia-based VoLTE network architecture deployed at Industrial Technology Research Institute (ITRI) in Taiwan. A UE (Fig. 1 (1)) accesses the evolved packet core (EPC) network (Fig. 1b) through an LTE base station called evolved node B (eNB, Fig. 1 (2)) in the Evolved Universal Terrestrial Radio Access Network (EUTRAN, Fig. 1a). The LTE spectrum considered in this article is Band 3 (1800 MHz). The eNB connects to the LTE core network with the mobility management entity (MME in Fig. 1 (3)) to send control messages through the S1-MME interface, and with the serving gateway (S-GW in Fig. 1 (4)) to send user data packets through the S1-U interface. In the LTE core network, the home subscriber server (HSS in Fig. 1 (5)) is the master database containing all user-related subscription information. The MME interacts with the HSS via the S6a interface to handle mobility management and session management of UE. The S-GW sends packets to external networks through the packet data network gateway (P-GW in Fig. 1 (6)) via the S5 interface. The Policy and Charging Rules Function (PCRF in Fig. 1 (7)) connects the P-GW via the Gx interface to determine the policy rules for the sessions established in the P-GW.

The LTE network connects to the IMS network (Fig. 1 (c)) through the SGi-IMS interface between the P-GW and the Proxy Call Session Control Function (P-CSCF in Fig. 1 (10)). The P-CSCF receives the signals from other networks and sends them to the Serving/Interrogating Call Session Control Function (S/I-CSCF Fig. 1 (9)) to handle the service requests through the Mw interface. The S/I-CSCF retransmits them to the telephony application servers (TAS in Fig. 1 (8)) via the ISC interface. The TAS is an application server that supports multimedia telephony service. The IMS network at ITRI is deployed with a Nokia CFX-5000 as the P-CSCF and the I/S-CSCF, and a Nokia Open TAS as the TAS. ITRI develops an E.164 number mapping (ENUM, Fig. 1 (11)) server connecting to the S/I-CSCF via the domain name service (DNS) protocol to translate the telephone numbers to Internet Protocol (IP) addresses.

To make a voice call, the UE sends a VoLTE request to the IMS network by using the Session Initiation Protocol (SIP). The IMS network receives the request through the P-CSCF, and negotiates with the P-GW to allocate the resources and establish the bearer for the VoLTE call. LTE also supports high data rate and low latency with the all-IP network to provide high-quality multimedia services such as video over LTE (ViLTE). The procedure to establish a video call is basically the same as that for a VoLTE call, where the P-GW and the S-GW establish two bearers for voice and the video, respectively [1].

In most commercial mobile operations, both 3G and LTE networks usually co-exist. LTE is a packet-switched (PS) network, where the circuit-switched (CS) voice service is not supported as it is in 3G. Typically, 3G provides more comprehensive radio coverage than the LTE coverage even if LTE voice service has been commercially deployed. Circuit switched fallback (CSFB) based on 3G [3] is still used to provide voice services by major operators in China and Taiwan. Also, single radio voice call continuity (SRVCC) is suggested to support voice service to transfer the LTE's IMS connection to the 3G CS connection when the LTE network is not available.

The 3GPP defines two modes of spectrum utilization for LTE: Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD). The FDD mode is used for symmetric uplink and downlink data rate. The TDD mode, on the other hand, supports asymmetric uplink and downlink

Yingrong Coral Sung is with National Chiao Tung University; Yung-Sheng Ho is with Industrial Technology Research Institute, Yi-Bing Lin and Jyh-Cheng Chen are with National Chiao Tung University; Herman Chung-Hwa Rao is with Far EasTone Telecommunications Co., Ltd.

Digital Object Identifier: 10.1109/MWC.2018.1700109

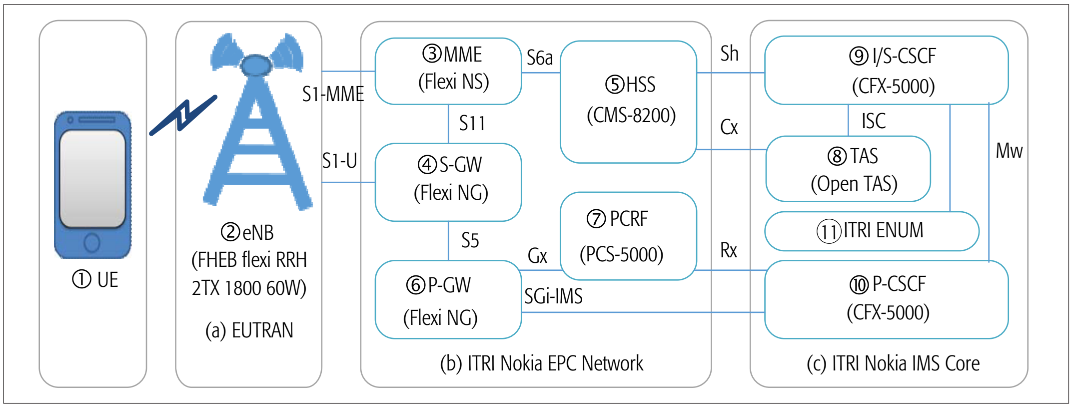

FIGURE 1. The ITRI Nokia network: a) EUTRAN; b) ITRI Nokia EPC Network; c) ITRI Nokia IMS Core.

The 3GPP defines two modes of spectrum utilization for LTE: Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD). The FDD mode is used for symmetric uplink and downlink data rate. The TDD mode, on the other hand, supports asymmetric uplink and downlink data rates with different uplink-downlink configurations.

data rates with different uplink-downlink configurations. The FDD mode and the TDD mode operate on different radio bands with different frame structures. Therefore, interworking between LTE FDD and LTE TDD networks is important for operators that have spectrums in both modes. In this article, we also present a multimode/multiband interworking environment established at Far EasTone (FET), a major operator in Taiwan. A comparison of these two modes can be found in [4].

Not many studies have reported the VoLTE and ViLTE measurement results and the testing environments, especially from real equipment or commercial networks. In this article, we introduce the ITRI VoLTE testbeds that support both laboratory and field tests. ITRI has deployed two LTE mobile networks for field tests: a Nokia VoLTE network (Fig. 1) and an Ericsson LTE network (Fig. 2). We also present the results of FET's experiments on multimode/multiband interworking ViLTE calls. FET has established the first multimode/multiband interworking ViLTE call in Taiwan. In this article, we conduct measurements to show how to test the qualities of VoLTE and ViLTE services. The empirical results serve as a good reference for operators to develop VoLTE and ViLTE services.

This article is organized as follows. The following section elaborates on the ITRI testbeds and demonstrates how to use the testbeds for VoLTE testing. Then we present the FET multimode/multiband interworking ViLTE experiments, which are based on an extension of the commercial core network described previously. The final section concludes this article.

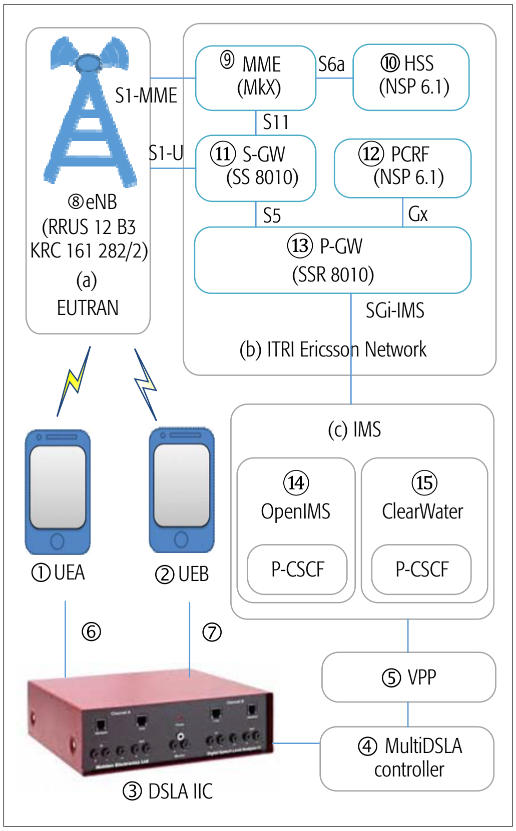

FIGURE 2. The ITRI Ericsson network and voice quality testing environment: a) EUTRAN; b) ITRI Ericsson Network; c) IMS.

The ITRI VoLTE Testing Environment

In this section, we describe the ITRI VoLTE testbeds supporting both laboratory and field tests. The laboratory utilizes R&S CMW500 as the test equipment to conduct functional and conformance tests. The functional tests include IMS registration, authentication, call establishment, termination, and so on. The conformance tests include IMS-CC (call control) [5], MTSI (multimedia telephony service for IMS), and short message service (SMS) over IMS. ITRI has deployed two mobile networks for field tests: a Nokia VoLTE network and an Ericsson LTE network with OpenIMS and Clearwater open source IMS solutions.

The ITRI Nokia VoLTE network (Fig. 1) supports VoLTE interoperability tests, including IMS registration, authentication, call establishment and termination, video call, CSFB and SRVCC, and supplementary services such as communication diversion, communication hold, ad-hoc multi-party audio conference, and communication barring. The eNB is a Nokia FHEB flexi RRH 2TX with 1800MHz band and 60 W power. The MME is a Nokia Flexi NS. Both S-GW and P-GW functionalities are implemented in a Nokia Flexi NG. The HSS is a Nokia CMS-8200, and the PCRF is a Nokia PCS-5000. The core network architecture is applicable to future 5G networks. The ITRI Ericsson network (see Fig. 2) is similar to the Nokia network except that it uses open-source IMS systems including OpenIMS (Fig. 2 (14)) and Clearwater (Fig. 2 (15)).

FIGURE 3. Functional blocks of ITRI VoLTE UE and its GUIs.

Testing Methodology

Based on Perceptual Evaluation of Speech Quality (PESQ) and Perceptual Objective Listening Quality Analysis (POLQA) [6], we use the ITRI VoLTE testbeds to evaluate the voice qualities for narrowband and wideband audio codecs. PESQ and POLQA evaluate the audio quality and output the MOS value by comparing the original audio data and the degraded audio data [7]. PESQ supports wideband voice with a model different from that for the voice frequency band. POLQA is an enhancement of PESQ. In addition to the voice frequency band ranging from approximately 300 Hz to 3400 Hz supported in PESQ, POLQA also supports wideband and super-wideband voice frequency band from 50 Hz to 14000 Hz. The scores for PESQ range from 1 to 5. On the other hand, the scores for POLQA-NB range from 1 to 4.5, and the scores for POLQA-SWB range from 1 to 4.75. The higher the score, the better the quality. The Malden Digital Speech Level Analyzer (DSLA) IIC (Fig. 2 (3)) and MultiDSLA Controller (Fig. 2 (4)) are used for speech quality measurements, which support two types of voice quality evaluation for the devices under test (DUT). The DUT-to-DUT scenario involves two UEs (Fig. 2 (1) and (2)) in the voice quality test. The DUT-to-VPP scenario involves one UE (Fig. 2 (1)) and a virtual UE VoxPort packet (VPP; Fig. 2 (5)). The VPP is an IMS end-point with VoIP signal generation and measurement capabilities within the IP domain. We use the Malden MultiDSLA controller to set up testing parameters and procedures. Based on these setups, DSLA IIC generates the source audio signal (reference signal) connected to the talker DUT through a headset port (. 2 (6)). The generated signal is processed at the DUT, delivered through the Ericsson or the Nokia networks, and then received and processed by the listener DUT. This DUT sends the processed signal (degraded audio signal) to the DSLA IIC through a headset port (Fig. 2 (7)). The DSLA IIC compares the reference signal (talker side) and the degraded signal (listener side) and then gives a voice quality report.

A DUT in Fig. 2 can be replaced by a VoLTE UE developed by ITRI, which is any LTE smartphone installed with the ITRI VoLTE software. The ITRI VoLTE UE is a terminal testing tool that can be set up or programmed to test specific features of other commercial UEs. This UE testing tool supports TCP/UDP for IMS signaling (Fig. 3 (1)) and Real-time Transport Protocol (RTP) for audio streaming (Fig. 3 (2)). The call control module (Fig. 3 (3)) is responsible for IMS registration, authentication with IMS Authentication and Key Agreement (AKAv1) or the MD5 Algorithm, subscription to the registration event package, mobile origination/termination calls, reliable provisional responses (100rel), provisional response acknowledgments (PRACK), and communication hold. The audio management module (Fig. 3 (4)) dynamically adjusts the AMR-NB mode. The audio codec module (Fig. 3 (5)) supports G.711U, G.711A, G.729, iLBC, and AMR-NB codecs. The adaptive multi-rate (AMR) audio codec is an audio compression format optimized for speech coding [8]. The AMR-NB codec supports eight modes with different bit rates: 4.75 kb/s (Mode 0), 5.15 kb/s (Mode 1), 5.90 kb/s (Mode 2), 6.70 kb/s (Mode 3), 7.40 kb/s (Mode 4), 7.95 kb/s (Mode 5), 10.2 kb/s (Mode 6), and 12.2 kb/s (Mode 7). The software architecture of the ITRI VoLTE UE flexibly supports module placement to adapt the UE to various test scenarios.

We have designed eight testing scenarios for POLQA measurements in the Nokia VoLTE network. For discussion purposes, these scenarios are grouped into three experimental sets. For each scenario, we measure the listening quality (at the receiver side) of both call parties 80 times, with 40 times from the calling party to the called party, and another 40 times from the called party to the calling party. We record the degraded signals of these 80 measurements respectively and calculate the average POLQA scores.

ARM-NB Mode 7 Tests with Same Types of UEs

This set of experiments tests the qualities of ARM-NB Mode 7 (12.2 kb/s). In Scenario I, both call parties are UEs of vendor A (UE A). In Scenario 2, both call parties are UEs of vendor B (UE B). In Scenario 3, both call parties are ITRI VoLTE UEs. In each scenario, the talker and the listener are UEs of the same type. Figure 4a shows the testing results. In Scenario 1, the POLQA scores of UE A range approximately from 3.1 to 3.9, and the average is 3.5. In Scenario 2, the scores of UE B range from 3.0 to 3.9, and the average is 3.5. In Scenario 3, the scores of ITRI VoLTE UE range from 2.6 to 3.9, and the average is 3.4. The measurement results indicate that the average scores for the three scenarios are the same. The variances of the commercial UEs are smaller than that for the ITRI UE. The reason is that ITRI UE is designed for testing and its software is implemented for engineering development with flexibility, which may or may not be intended for optimized commercial usage.

ARM-NB Mode 7 Tests with Different Types of UEs

This set of experiments tests the qualities of ARM-NB Mode 7 where the talker and the listener are different types of UEs. In Scenario 4, the call parties are UE A and UE B. In Scenario 5, the call parties are UE B and ITRI UE. Figure 4b shows the POLQA scores of VoLTE calls between UE A and UE B for Scenario 4, where the scores measured at UE A (as the listener) range from 3 to 3.9, and the average is 3.6. The scores measured at UE B range from 3.3 to 3.9, and the average is 3.6. An interesting observation is that the POLQA scores in this scenario are better than those for Scenarios 1–3. That is, when the call parties are the UEs of different brands, the voice quality is better than that for the UEs of the same brands.

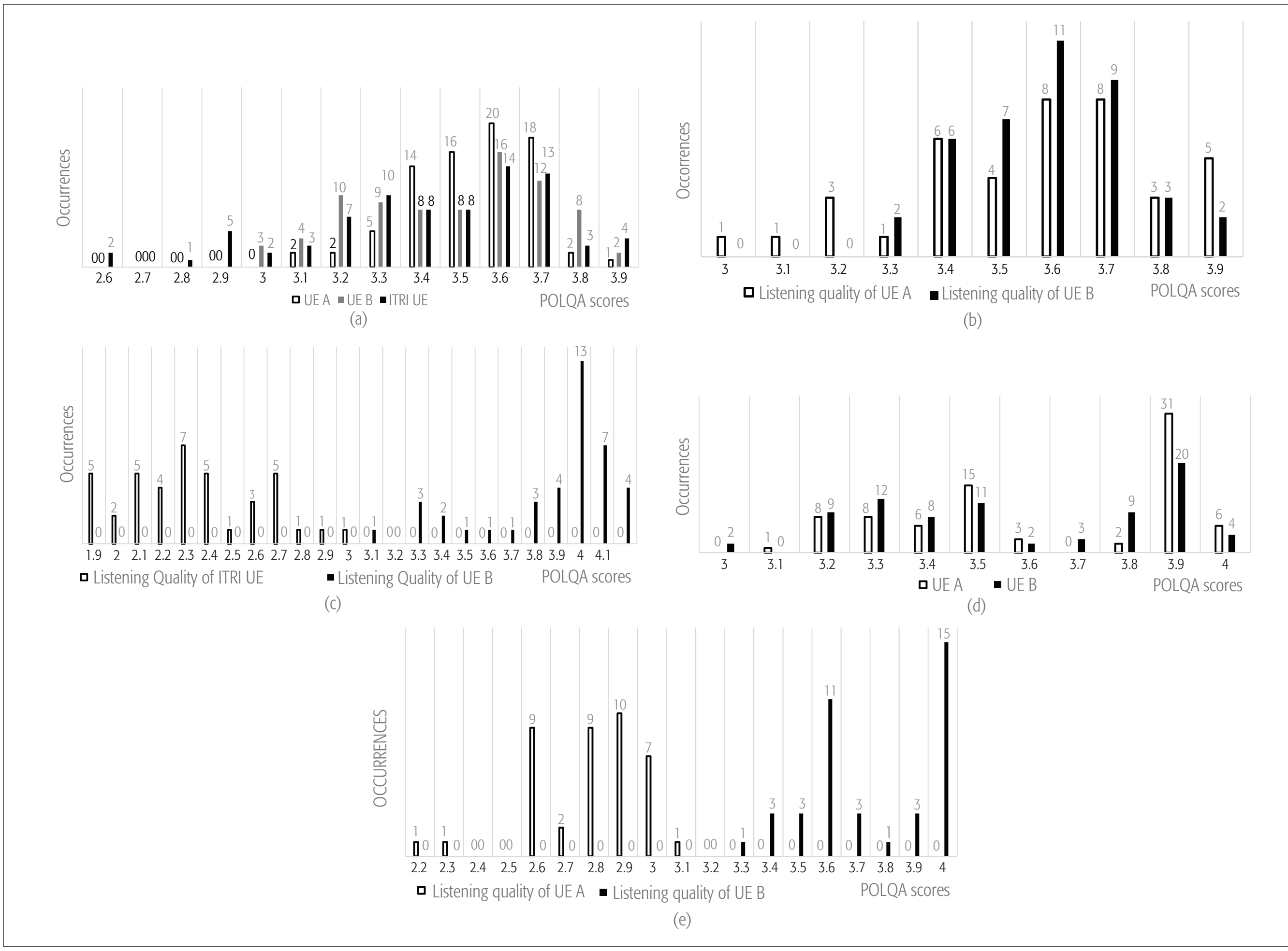

FIGURE 4. POLQA scores for eight scenarios: a) scenarios 1-3: AMR-NB between UEs; b) scenario 4: AMR-NB between UE A and UE B; c) scenario 5: AMR-NB between UE B and ITRI UE; d) scenarios 6 and 7: AMR-WB between UEs; e) scenario 8: AMR-WB between UE A and UE B.

Figure 4c shows the POLQA scores where the call parties are UE B and ITRI UE in Scenario 5. The scores measured at UE B range from 3.1 to 4.2, and the average is 3.9. The scores measured at the ITRI UE range from 1.9 to 3, and the average is 2.3. The results indicate that by pairing with the ITRI UE, the voice quality of UE B is significantly enhanced. In contrast, the voice quality of the ITRI UE is significantly degraded. We suggest that the narrowband voice decoder of ITRI UE should be improved for better performance for the following reason. Both the voice encoder and the decoder contribute to the audio quality. In the past, handset manufacturers were asked to modify both the encoder and the decoder of different brands to enhance the audio quality. The results were not good. Tuning of the encoder depends on the results produced by the decoder. After the encoder is tuned, we have to send the packets of the modified encoder to the decoder to generate the new results for the next adjustment of the encoder. This process involves both the encoder and the decoder, and has to repeat iteratively. We suggest that for the UEs of the same brand, tune both the encoder and the decoder, which is typically done before the UE is commercially released. For different brands of UEs, we suggest using the audio traces generated by the encoder to adjust the decoder. Every time the decoder is adjusted, we can use the same audio trace to test the modified decoder, and there is no need to involve the encoder. The audio traces generated by various brands of UEs can be saved in a database and used in parallel to tune a decoder. In this way, we can save the audio traces encoded by various encoders in a database, and use them to tune a decoder.

ARM-WB Mode 8 Tests with Commercial UEs

This set of experiments tests the qualities of ARM-WB Mode 8 (23.85 kb/s) for two UE As in Scenario 6, two UE Bs in Scenario 7, and both UE A and UE B in Scenario 8. In Scenarios 6 and 7, the talker and the listener are UEs of the same type. Figure 4d shows that in Scenario 6, the scores of UE A range from 3.1 to 4.0, and the average is 3.6. In Scenario 7, the scores of UE B range from 3.0 to 4.0, and the average is 3.6. Compared with Scenarios 1-3 (with ARM-NB Mode 7), Scenarios 6 and 7 (with ARM-WB Mode 8) have better voice qualities with smaller variances (more stable).

Figure 4e shows the POLQA scores of VoLTE calls between UE A and UE B. The scores measured at UE A range from 2.2 to 3.1, and the average is 2.8. The scores measured at UE B range from 3.0 to 4.0, and the average is 3.7. In this Scenario, we observe the same phenomenon in

ARM	Scenario	Call parties	POLQA score
ARM-NB mode 7	Scenario 1	Two UE A (vendor A)	Range [3.1, 3.9]; the average is 3.54.
	Scenario 2	Two UE B (vendor B)	Range [3.0, 3.9]; the average is 3.48.
	Scenario 3	Two ITRI VoLTE UEs	Range [2.6, 3.9]; the average is 3.42.
	Scenario 4	UE A and UE B	UE A range [3, 3.9]; the average is 3.57. UE B range [3.3, 3.9]; the average is 3.59.
	Scenario 5	UE B and ITRI UE	UE B range [3.1, 4.2]; the average is 3.88. ITRI UE range [1.9, 3]; the average is 2.34.
ARM-WB mode 8	Scenario 6	Two UE A	Range [3.1, 4.0]; the average is 3.64.
	Scenario 7	Two UE B	Range [3.0, 4.0]; the average is 3.58.
	Scenario 8	UE A and UE B	UE A range [2.2, 3.1]; the average is 2.79. UE B range [3.0, 4.0]; the average is 3.76.

TABLE 1. Summary of eight scenarios.

Scenario 5. That is, by pairing a UE A with a UE B, the voice quality of UE B is significantly enhanced while the voice quality of UE A is significantly degraded. This non-trivial result suggests that the wideband voice decoder of UE A needs to be improved. Details of the above eight scenarios are summarized in Table I.

The Impact of the Background Traffic

We have measured the POLQA scores for scenario 1 (vendor A's UEs) and scenario 2 (vendor B's UE) with and without background traffic. The background traffic is generated with the TCP packets using iperf and the data rate is 14.1 Mb/s, which is the bearer capacity of the data traffic allocated for the UEs. The histograms for the measures with and without background traffic are different. However, in terms of traffic engineering, the most important metrics, that is, the mean and the standard deviations for the POLQA scores, are about the same with or without background traffic. In most cases, the POLQA scores are not affected significantly by the background traffic because the bearers used for data traffic and for IMS voice streaming are separated. For vendor A's UE, the average POLQA score is 3.47 and the standard deviation is 0.24 with background traffic. Without background traffic, the average POLQA score is 3.50 and the standard deviation is 0.26. For vendor B's UE, the average POLQA score is 3.22 and the standard deviation is 0.18 with background traffic. Without background traffic, the average POLQA score is 3.24 and the standard deviation is 0.16. Both cases show that the background traffic does make the difference. However, the POLQA sore without background traffic is only slightly better than that with background traffic.

Comparison of the Testing Environments

We compare several related studies with the ITRI testing environment. The quality of VoLTE calls is quantified by mean opinion score (MOS), which rates the audio quality from 1 (bad) to 5 (excellent) [6]. MOS accounts or penalizes the environment's effect in our experiments through packet loss and transmission delay. In some related studies, the MOS values are obtained from humans listening to the audio and rating the audio quality, or transforming the results of other algorithms such as POLQA, PESQ, and so on.

Table 2 summarizes the conditions of several testing environments that measure the quality of the VoLTE services using the POLQA or the WB-E models [9] under different signal strengths in terms of reference signal received power (RSRP) [10]. Y. J. Jia *et al*. [11] measured the quality of VoLTE services in the field environment and reported that the mean MOS value is 3.8 and higher under good signal strength levels. The audio codec used is not specified in this work. M. D. Villaluz *et al*. [12] conducted the VoLTE quality measurement experiments in a field environment in a wide signal strength range and obtained MOS values from 1.19 to 4.15. The above two studies did not specify which commercial core networks are used. The type of UEs and used audio codec are not reported in the work. D.-H. Nguyen *et al*. [13] evaluated the quality of VoLTE service with lab simulations and reported that the MOS values are higher than 4 in most cases. In comparison with the above previous works, our work in the ITRI VoLTE testing environment reports more details of the experiments. We measure the quality of VoLTE services in the field environment with both commercial devices and ITRI-developed UE to provide more references to the VoLTE performance.

Multimode and Multiband Interworking

In this section, we evaluate the quality of ViLTE service in multimode and multiband interworking scenarios established in the commercial FET environment, an extension of the core networks

	Y. J. Jia *et al*., [11]	M. D. Villaluz *et al*., [12]	D.-H. Nguyen *et al*., [13]	ITRI Tests
Type of UEs	Commercial devices	N/A	Self-developed UE	Both commercial devices and self-developed UE
Core network	N/A	N/A	Lab simulation	Nokia, Ericsson
Signal strength (RSRP)	[–95 dBm, –130 dBm]	[–75 dBm, –120 dBm]	N/A	[–95 dBm, –110 dBm]
Audio codec	N/A	N/A	AMR-WB	AMR-NB and AMR-WB
Algorithm	POLQA	POLQA	WB E-model	POLQA
Reported MOS value range	[1.5, 4.2]	[1.19, 4.16]	[3.47, 4.47]	[2.2, 4.2]

TABLE 2. Comparison of VoLTE tests.

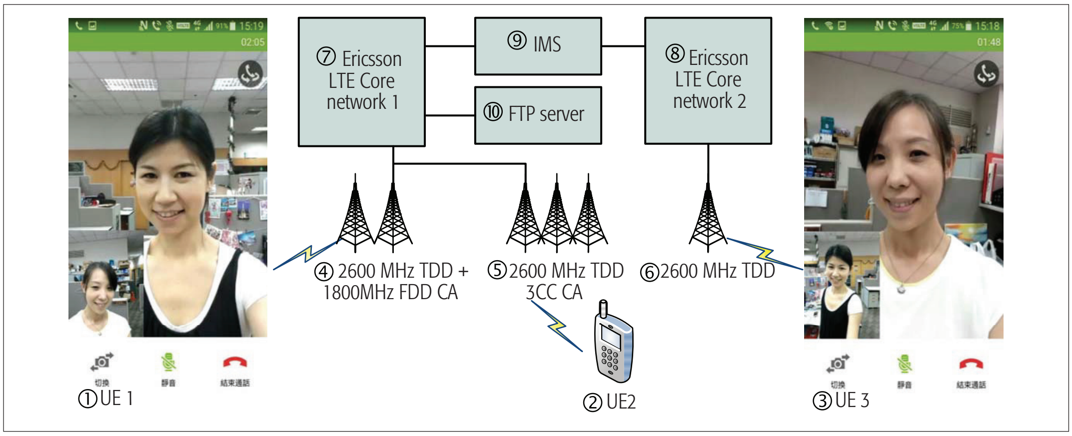

FIGURE 5. Multimode/Multiband interworking architecture.

MOS accounts or penalizes the environment's effect in our experiments through packet loss and transmission delay. In some related studies, the MOS values are obtained from humans listening to the audio and rating the audio quality, or transforming the results of other algorithms such as POLQA, PESQ, and so on.

illustrated in Fig. 2. Figure 5 illustrates the interworking architecture that consists of both the FDD and the TDD eNBs.

In this trial, each of the UEs (Fig. 5 (1), (2), and (3)) is an HTC One M9 with the Qualcomm snapdragon chipset. These UEs support carrier aggregation (CA) to combine LTE multimode and multiband radio frequencies for better radio performance. HTC One M9 supports the FDD mode operating on Band 3 (i.e., uplink band at 1710 – 1785 MHz and downlink band at 1805 – 1880 MHz) and Band 28 (i.e., uplink band at 703 – 748 MHz and downlink band at 758 – 803 MHz). In addition, it supports the TDD mode operating on Band 41 (i.e., operate band at 2496 – 2690 MHz). In FET's network, the eNBs (Fig. 5 (4), (5), and (6)), Ericsson LTE Core Networks (Fig. 5 (7) and (8)), and the IMS network (Fig. 5 (9)) are built with Ericsson commercial products. The model type of eNBs is RBS 6601 DUS 41, which supports FDD mode operating on Band 28. The eNB connects radio unit type RRUS 61 B41C for TDD mode and RRUS 12 B3 for FDD mode. The links between the eNBs and LTE core networks are 1Gb/s optical fiber cables. Table 3 summarizes the multimode and multiband interworking experiments conducted in the commercial FET environment.

We consider two scenarios to investigate throughput performance. In the first scenario, the eNBs are configured to use CA to combine Band 38 TDD (2600MHz) with 20 MHz bandwidth and Band 3 FDD (1800MHz) with 15 MHz bandwidth (Fig. 5 (4)). The antenna configuration is 22 MIMO and the MCS is 28. UE 1 (Fig. 5 (1)) connects to the Ericsson LTE core network 1 (Fig. 5 (7)) through the eNBs to download streaming files from an FTP server (Fig. 5 (10)) in the external network. The throughput is 231 Mb/s. In the second scenario, the eNBs are configured to use Band 41 TDD (2600MHz) with 3 Component Carrier Aggregation (3CC CA), that combines three 20 MHz bandwidths. UE 2 (Fig. 5 (2)) connects to the LTE core network 1 to download streaming files from the FTP server in the external network. The throughput is 369.86 Mb/s. The throughput performance is good enough to support ViLTE services.

To perform the multimode/multiband interworking ViLTE calls, UE 1 operates in Band 28 uplink band at 713 – 723 MHz and downlink band at 768 – 778MHz, and UE 3 (Fig. 5 (3)) operates in Band 41 (2635 – 2655 MHz). Table 4 shows the video call performance in the 2G/3G environment, 3G/LTE environment with CSFB, and LTE environment with ViLTE service. The MOS values are obtained from a SAGE 960B, which uses the PSQM algorithm [6] to evaluate the audio quality. In our LTE environment, the ViLTE call setup time is about two seconds, which is two times faster than in 2G/3G and CSFB. The result is due to the fact that the ViLTE call setup does not involve heterogeneous telecommunication systems such as 2G and 3G. In a pure LTE environment, the video call provides better voice/video quality. The voice is HD AMR-WB quality with bit rate 12.65 kb/s, and the video is SVGA quality with bit rate 1.4 Mb/s. The MOS value of the ViLTE call in the LTE environment is 4.08,

Experiment	UE configuration	Bandwidth
FTP Scenario 1	Band 3 FDD and Band 38 TDD CA	35 MHz
FTP Scenario 2	Band 41 TDD with 3CC CA	60 MHZ
ViLTE call	Band 28 FDD	20 MHz
	Band 41 TDD	

TABLE 3. Summary of the multimode and multiband interworking experiments.

	2G/3G	CSFB	ViLTE
Data/Voice	CS/PS	CS/PS	PS/PS
Call setup time	~ 5 sec.	~ 6 – 8 sec.	~ 2 sec.
Voice quality	AMR-NB	AMR-NB	HD AMR-WB
Video quality	64 CSD	64 CSD	SVGA
MOS	3.6	3.6	4.08

CS: circuit switched; PS: packet switched; AMR-NB: adaptive multi-rate narrowband; HD: high definition; AMR-WB: adaptive multi-rate wideband; CSD: circuit switched data; SVGA: super video graphics array

TABLE 4. The performance of video calls in different systems.

We provided first-hand experience on the ITRI testing environment for many commercial handset products, and a reference for operators to develop their VoLTE and ViLTE services, which are appreciated by both UE manufacturers and mobile operators. The results are also valuable for a general reader who wants to know about the VoLTE/ViLTE performance testing with the commercial core network.

which is better than that in 2G/3G and CSFB. The screen shots of UE 1 and UE 3 (Fig. 5 (1) and (3)) show that the video quality is good without any broken blocks. The experiments indicate that multimode and multiband interworking ViLTE call quality conforms with the QoS classes specified in 3GPP TS 36.300 [14]. Note that the ViLTE call experiments can also be conducted with the first configuration. Our experiments show that, with or without CA, the results for the ViLTE call's performance are the same.

Conclusion

In this article, we present the VoLTE and ViLTE testing environments. We investigate VoLTE quality tests in ITRI testbeds with commercial equipment. We also compare several related studies with the ITRI testbeds. We report more details of the experiments than previous works and provide more references to the quality measurement of VoLTE calls.

We also present an LTE multimode and multiband interworking ViLTE call with a real trial. The call setup time of the ViLTE call is two times faster than in 2G/3G or CSFB video call. Also, the MOS value of the call achieves 4.08. The experimental results indicate that the call quality of multimode and multiband interworking ViLTE conforms with the QoS classes specified in 3GPP TS 36.300.

We summarize the results of VoLTE/ViLTE quality measurements as follows:

- When the call parties are the UEs of different brands, the voice quality of ARM-NB Mode 7 is better than that for the UEs of the same brands.
- The performance of ARM-WB Mode 8 is affected by the UEs of different brands, and thus it is suggested to conduct the call measurements by paring UEs of various brands.
- The VoLTE call quality is evaluated based on POLQA-NB and POLQA-WB to provide more general results. The POLQA-WB scores are higher than the POLQA-NB scores on average.
- The UEs may have bad performance due to the matching problem of encoders and decoders.
- In our LTE environment, the ViLTE call-setup time is about two seconds, which is two times faster than that in 2G/3G and CSFB. This result is due to the fact that the ViLTE call setup does not involve heterogeneous telecommunication systems such as 2G and 3G.
- In comparison with 2G/3G and CSFB system, a video call in the pure LTE environment provides better voice quality and video quality.

Our observation guides the direction for mobile operators to instruct the handset manufacturers to analyze their products. Such guidance is valuable for handset testing. According to 3GPP [2], VoLTE/ViLTE will be used in 5G systems. Therefore, the testing methodology described in this article can also be used in future 5G systems. Right now the 5G new radio based on 3GPP has not been deployed yet. When the 5G radio is available in the future, we will use the same testing procedure to evaluate the VoLTE/ViLTE performance.

In summary, we have conducted VoLTE and ViLTE quality measurement experiments with commercial equipment and present the details of the experiment setups. We have enhanced the test setup and metric that has not been reported in the literature. We provided first-hand experience on the ITRI testing environment for many commercial handset products, and a reference for operators to develop their VoLTE and ViLTE services, which are appreciated by both UE manufacturers and mobile operators. The results are also valuable for a general reader who wants to know about the VoLTE/ViLTE performance testing with the commercial core network.

Acknowledgement

This work is supported in part by MOST 106-2221-E-009 -049-MY2, MOST 106-2218-E-009-029, MOST 106-2622-8-009-017, MOST 106-2218-E-009-016 and 106T0D5096108. This work is also sponsored in part by ITRI.

References

[1] 3GPP; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2; Technical Specification 3GPP TS 23.228 version 13.3.0 (2015-06), 2015.

[2] 3GPP on track to 5G, http://www.3gpp.org/news-events/3gpp-news/1787-ontrack_5g

[3] 3GPP; Technical Specification Group Services and System Aspects; Circuit Switched (CS) fallback in Evolved Packet System (EPS); Stage 2; Technical Specification 3GPP TS 23.272 version 13.4.0 (2016-06), 2016.

[4] 3GPP; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, Technical Specification 3GPP TS 36.101 version 13.0.0 (2015-07), 2015.

[5] 3GPP; Technical Specification Group Services and System Aspects; Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); User Equipment (UE) conformance specification; Part 3: Abstract test suite (ATS); Technical Specification 3GPP TS 34.229-3 version 13.1.0 (2017-01), 2017.

[6] ITU-T; Perceptual objective listening quality assessment; ITU-T Recommendation P.863, 2014.

[7] Y.-C. Sung *et al.*, "NCTU-VT: A Freeware for Wireless VoIP Performance Measurement," *Wireless Communications and Mobile Computing*, vol. 12, no. 4, 2012, pp. 318–24.

[8] 3GPP; Technical Specification Group Services and System Aspects; ANSI-C code for the Adaptive Multi Rate (AMR) speech codec; Technical Specification 3GPP TS 26.073 version 14.0.0 (2017-03), 2017.

[9] ITU-T; Wideband E-model; ITU-T Recommendation G.107.1, 2015.

[10] 3GPP; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements, Technical Specification 3GPP TS 36.214 version 14.1.0 (2017-01), 2017.

[11] Y. Jack Jia *et al.*, "Performance Characterization and Call Reliability Diagnosis Support for Voice over LTE," *Proc. 21st Annual Int'l. Conf. Mobile Computing and Networking (MobiCom '15)*, 2015, pp. 452–63.

[12] M. D. Villaluz *et al.*, "VoLTE SRVCC Optimization as Interim Solution for LTE Networks with Coverage Discontinuity," *Proc. 2015 Int'l. Conf. Information and Communication Technology Convergence (ICTC)*, 2015.

[13] D.-H. Nguyen and H. Nguyen, "A Dynamic Rate Adaptation Algorithm Using WB E-Model for Voice Traffic over LTE Network," 2016 Wireless Days (WD), 2016.

[14] 3GPP; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, Technical Specification 3GPP TS 36.300 version 14.1.0 (2016-12), 2016.

Biographies

Yingrong Coral Sung received the BSCSIE, MSCSIE, and Ph.D. degrees from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 2002, 2003, and 2010 respectively. She is currently an associate research fellow at the Microelectronics and Information Research Center at NCTU. Her current research interests include the design and analysis of personal communications services networks, network security and performance modeling.

Yung-Sheng He received his master degree in computer science from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 2003. He is now a senior software engineer and manager of the Broadband Networks and System Integration Technology division, Information and Communications Research Lab (ICL) at the Industrial Technology Research Institute (ITRI). His research interests include Internet telephony, virtualization technology, and multi-access edge computing.

Yi-Bing Lin [M'96, SM'96, F'03] received his Ph.D. from the University of Washington, USA, in 1990. From 1990 to 1995 he was a research scientist with Bellcore (Telcordia). He then joined National Chiao Tung University (NCTU) in Taiwan, where he remains. In 2010, he became a lifetime Chair Professor at NCTU, and in 2011, the vice president of NCTU. From 2014 to 2016, he was Deputy Minister, Ministry of Science and Technology, Taiwan. Since 2016, he has been serving as Vice Chancellor of the University System of Taiwan. He is an adjunct research fellow, Institute of Information Science, Academia Sinica, Research Center for Information Technology Innovation, Academia Sinica, and a member of the board of directors, Chunghwa Telecom. He is the author of the books *Wireless and Mobile Network Architecture* (Wiley, 2001), *Wireless and Mobile All-IP Networks* (John Wiley,2005), and *Charging for Mobile All-IP Telecommunications* (Wiley, 2008). He is Chair of IEEE Taipei Section. He is AAAS Fellow, ACM Fellow, IEEE Fellow, and IET Fellow.

Jyh-Cheng Chen [S'96, M'99, SM'04, F'12] has been a faculty member with National Chiao Tung University (NCTU), Hsinchu, Taiwan since 2010. Prior to that, he was with Bellcore/Telcordia Technologies in New Jersey, USA, and National Tsing Hua University (NTHU), Hsinchu, Taiwan. He is also now serving as the Convener, Computer Science Program, Ministry of Science and Technology, Taiwan. He has received numerous awards, including Outstanding Teaching Awards from both NCTU and NTHU, the Outstanding Research Award from the Ministry of Science and Technology, the Outstanding I. T. Elite Award, Taiwan, the K. T. Li Breakthrough Award from the Institute of Information and Computing Machinery, and the Telcordia CEO Award. He is a Fellow of the IEEE and a Distinguished Member of the ACM. He was a member of the Fellows Evaluation Committee, IEEE Computer Society.

Herman Rao is currently the Executive Vice President of the Network & Technology Division at Far EasTone Telecommunications Co., Ltd. in Taiwan. He is responsible for the development and operation of mobile and broadband networks and ISP. He joined FET in 1997 as a Director responsible for Technology Research & Development, Network Architecture, Service Platform & Deployment. In 2000, he was appointed Vice President. He also headed the planning of FEG Tpark as COO. He has extensive industry experience in telecommunications and data communications. Prior to FET, he worked at Bell Labs as a senior researcher and AWS AT&T Wireless as a research director for 10 years. He has ten U.S. patents. He was also an associate professor with the National Tsing-Hua University and National Chung-Cheng University. He holds a doctorate in computer science from the University of Arizona and a bachelor of science in mechanical engineering from the National Taiwan University.