(Article begins on next page)

A Comparative Study on the Quality of Narrow-Band and Wide-Band AMR VoLTE Calls

Elena Cipressi*† and Maria Luisa Merani†
*Empirix, Modena, Italy, Email: ecipressi@empirix.com
†Dipartimento di Ingegneria "Enzo Ferrari", University of Modena and Reggio Emilia, Modena, Italy
Email: [elena.cipressi, marialuisa.merani]@unimore.it

Abstract—This work performs a comparative analysis of the end-to-end quality guaranteed by Voice over LTE (VoLTE), examining several millions of VoLTE calls that employ two popular speech audio codecs, namely, Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wide Band (AMR-WB). To assess call quality, VQmon®, an enhanced version of the standardized E-Model, is utilized. The study reveals to what extent AMR-WB based calls are more robust against network impairments than their narrowband counterparts; it further shows that the dependence of call quality on the packet loss rate is approximately exponential for both types of codec.

Keywords—VoLTE; Quality of Experience; LTE Network Measurements.

I. INTRODUCTION

A. Rationale and Contribution

Long Term Evolution (LTE) and LTE-Advanced (LTE-A) are the latest standardized technologies for cellular connectivity and their evolution poses new basis to the commercial advent of the fifth generation of networks (5G). As a matter of fact, 5G standardization will deeply leverage on the progresses of several LTE features and services, such Voice over LTE (VoLTE). Solutions for supporting voice services in LTE have been historically built on two distinct technologies: (i) Circuit Switched Fall Back (CSFB) [1], that relies on the preexistent GSM/UMTS networks; (ii) VoLTE via IP Multimedia Subsystem (IMS), defined by GSMA in 2014 [2], where voice functionality is provided by an architectural framework paired to the LTE core network. The main advantage of VoLTE via IMS is the exploitation of LTE architecture, with no dependency upon external GSM/UMTS networks. Moreover, the IMS is in charge of the interworking with legacy 2G/3G networks, thus supporting call continuity in case of LTE coverage losses.

On the market, the road to VoLTE is partially paved: in some countries VoLTE is experiencing widespread diffusion, whereas the adoption of such technology is in its early stages in many other regions. Main standardization bodies, such as the Third Generation Partnership Project (3GPP) and the International Telecommunications Union (ITU), are at the forefront to enhance and promote VoLTE deployment. The reason is two-fold: firstly, there is an acclaimed urgency to release the spectrum that is currently assigned to 2G/3G operators;

secondly, as 5G promises to shift the network paradigms from network-centricity to user-centricity, VoLTE, assisted by wide-band and super-wide band codecs, is the best candidate to perform high definition calls and to ensure high quality in a totally IP-based scenario.

In view of the forecast user centric scenario, this paper leverages on over ten million calls collected from a real LTE commercial network to assess the quality that a VoLTE user shall expect. In doing so, a network perspective is taken, focusing on the effects that different values of packet loss rate and maximum jitter have on the quality of voice calls. The analysis is centered on two widespread speech audio codecs, namely, Adaptive Multi Rate (AMR) and Adaptive Multirate Wide Band (AMR-WB). Several illuminating results are provided, that can be summarized as follows:

1) on a well designed LTE network, the packet loss rate and the maximum jitter that voice calls experience are confined and the network parameter that mostly influences their quality is the packet loss rate;

2) calls employing the AMR-WB codec are more robust against the packet loss rate: not only their quality is superior, but it also exhibits a lower standard deviation than the narrowband counterpart. Further, the maximum jitter experienced by AMR-WB calls has a very modest effect on quality;

3) the dependency of reconstructed voice quality on the packet loss rate is successfully captured by an exponential law for both narrowband and wideband speech audio codecs.

B. Related Work

In the past, several analysis have been conducted to evaluate the impact that network impairments have on the quality experienced by end-users for different type of services, such as real-time communications and streaming applications [3]-[4]. Within this framework, in [3] Fiedler et al. gave a major contribution by observing that *"generic Quality of Service (QoS) problems* (e.g., loss, delay, jitter, reordering) imply *generic Quality of Experience (QoE) problems* (e.g. glitches, artifacts, impairments of various kind)". Moreover, they expressed the functional dependency of QoE by QoS through a differential equation whose solution is an exponential function.

They successfully proved the mathematical foundation of their work for Skype-VoIP, a popular voice-call service affected by packet loss, jitter and reordering.

The authors of [5] exploited a test framework consisting of a UMTS simulator for the air interface and an IP network simulator for the transmission of the IP packets on the Core Network to perform real-time conversational tests. Their results showed that the AMR and AMR-WB speech codecs are well-suited for packet switched conversational applications. More recently, the performance of commercially deployed VoLTE was characterized by means of controlled experiments in [6]; in detail, a comparison was set up, to confront VoLTE against circuit-switched and Skype/Google Hangouts voice calls. In [7], the performance of VoLTE and of Circuit-Switched Fall Back was benchmarked, pinpointing what values of call set up delay can be achieved under various radio conditions. In [4], the dependency of the average VoLTE call duration on call quality was investigated. Finally, in [8] the authors' objective was to understand whether the adoption of a lower bit rate of the AMR-WB codec could result in an augmented coverage for VoLTE users.

Differently from previous contributions, the aim of this paper is to discern the dependency of VoLTE call quality on network impairments, i.e., packet loss and jitter, and to grasp the influence that different codec choices, namely, AMR or AMR-WB, have on end-to-end speech quality. These goals are achieved inspecting a real LTE network and accordingly examining a significantly large set of VoLTE calls: the network conditions they encountered were recorded and their quality estimated via VQmon®. The obtained results allow to realistically compare the behavior of AMR and WB-AMR codecs and to shed light on VoLTE performance.

The remainder of this paper is organized as follows. Section II depicts the background for this work, touching upon VoLTE architecture, currently adopted voice codecs and objective voice quality assessment. Section III illustrates the data collection process and then discusses the main measured characteristics of the network under examination, as well as the results obtained in terms of voice quality analysis. Last, Section IV draws the conclusions.

II. Background

This Section serves different purposes. It first illustrates VoLTE architecture, so as to understand the approach undertaken to monitor VoLTE calls. It next summarizes the most salient features of the AMR and AMR-WB voice codecs, that are the subject of the current investigation. It finally provides a brief description of the tool employed for objective voice quality evaluation.

A. VoLTE Architecture

Fig1 reports the main elements of the LTE network that are involved in a VoLTE call, along with the standard interfaces traversed by the data and signaling flows. Within the Evolved Universal Terrestrial Radio Access Network (E-UTRAN), they are: (i) the User Equipment (UE) of the subscriber engaged in the conversation and (ii) the e-NodeB (eNB), being responsible to allocate UEs radio resources on the uplink and downlink, as well as to protect the UE sensitive data crossing the Uu radio interface via a suitable encryption method.

Within the Evolved Packet Core (EPC), the Mobility Management Entity (MME), the Serving Gateway (SGW) and the Packet Data Network Gateway (PGW) are the next blocks encountered. The MME is the key access control element in LTE, as it is in charge of the proper SGW choice whenever a UE attaches to the network. The SGW forwards the user plane data packets to an eNB and/or to a PGW, that in turn takes care of the connection between the UE and the outside, e.g., the Internet. For a VoLTE call, the PGW connects to the IP Multimedia Subsystem (IMS), a VoIP platform whose main constituent elements are the Media Resource Function Processor (MRFP) and the Media Resource Function Controller (MRFC); the former handles the RTP packets, carrying voice samples, the latter takes care of the associated signaling, provided by the Session Initiation Protocol (SIP). The PGW and the IMS communicate via the Mb interface.

The E-UTRAN plus the EPC, i.e., the Evolved Packet system (EPS), is connection oriented; hence, after the UE has connected to the network and the authentication process has successfully come to an end, a first virtual connection, called Default EPS Bearer, is activated. In this circumstance, and differently from what happens in UMTS, the UE is assigned an IP address. Moreover, the QoS class assigned to this first bearer, which in the LTE jargon is the Quality Class Identifier[9], is set equal to 9 (i.e., the lowest priority), that corresponds to a packet loss rate equal to 10^{-6}, and to a delay budget of 300 ms, a combination deemed acceptable for non-Guaranteed Bit Rate (non-GBR) applications.

When a VoLTE call is performed, two additional logical pipes have to be opened between the UE and the network. Namely, a second Default Bearer is activated with the IMS network: it will be responsible for carrying the SIP signaling between the UE and the IMS. This bearer is assigned a QCI equal to 5, i.e., the highest priority level, requiring from the network a packet loss rate of 10^{-6}, and a packet delay budget of 100 ms, apt for GBR traffic. The third, dedicated bearer is finally activated for the delivery of the voice media packets, with $QCI = 1$, that corresponds to the second highest priority, that is, a packet loss rate of 10^{-2} and a stringent packet delay budget of 100 ms, fulfilling the needs of conversational voice.

B. Codec Overview

When it comes to the voice codecs most widely adopted in current days, we observe that the Adaptive Multi-Rate (AMR) [10] codec is a narrow-band codec largely popular in GSM and UMTS. It was originally developed by the European Telecommunication Standard Institute (ETSI) for GSM cellular systems and it was then chosen by 3GPP as a standard speech codec for UMTS, as it overcomes the limitations of the previous standardized GSM Enhanced Full-Rate (EFR) codec [11]. The AMR encoder is able to dynamically adapt its output rate to the current radio channel conditions, featuring
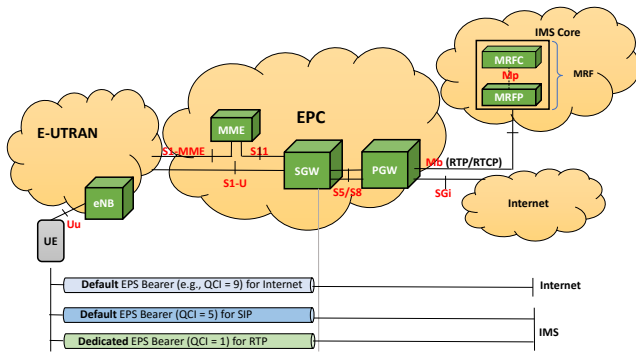
Fig. 1. Simplified LTE and IMS network architecture

R-factor	MOS	Speech Quality	User Satisfaction
≥ 90	≥ 4.34	Best	Very Satisfied
≥ 80	≥ 4.03	High	Satisfied
≥ 70	≥ 3.60	Medium	Some Users Dissatisfied
≥ 60	≥ 3.10	Low	Many Users Dissatisfied
≥ 50	≥ 2.58	Poor	Nearly All Users Dissatisfied

eight different source rates of 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2 kbit/s. The sampling frequency it uses is 8000 Hz and the duration of the speech frames it produces is 20 ms. Hence, each encoded AMR speech frame carries 160 samples of the original speech. User terminals that are VoLTE capable must be able to operate in all AMR eight modes.

In 1999, 3GPP together with ETSI began the standardization of a wide-band speech codec designed for WCDMA 3G and GSM systems. The process was finalized at the beginning of 2001, when the Adaptive Multi Rate WideBand codec (AMR-WB) [12] specifications were approved. Nowadays, AMR codec is going to be progressively replaced by its AMR-WB counterpart, capable of operating with nine source rates of 6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 and 23.85 kbit/s. The sampling frequency in AMR-WB is 16000 Hz and each encoded speech frame carries 320 samples of the original speech. If wide-band speech communication is offered as part of the VoLTE service, all nine modes must be supported by the user terminal. Whereas the AMR codec has been optimized for the voice components falling within the $[300, 3400]$ Hz frequency window, AMR-WB covers a wider frequency range, spanning from 50 Hz to 7000 Hz. Such broader bandwidth increases the intelligibility and the naturalness of the reconstructed speech, easing the recognition of the speaker. The official ITU-T test outcomes reported in [13] and [12] prove the substantial improvement of perceived voice quality provided by the bandwidth extension from narrowband to wideband. On the industry rim, in 2006 T-Mobile (Deutsche Telekom AG) in partnership with Ericsson, collected the results of subjective tests administered to a pool of 150 external research participants, with approximately 80% of them claiming to have *"heard distinct differences between normal and high voice quality call"*.

C. Non-intrusive Voice Quality Monitoring and VQmon®

Voice quality monitoring is a crucial topic for mobile operators, and as such has recently experienced an increased interest. The approaches to voice quality assessment are broadly classified in subjective and objective, the former mandating for a pool of listeners that rate the quality of test calls, the latter relying on automated algorithms. As subjective tests are costly, hard to repeat and time consuming for massively and periodically performed measurements campaigns, objective tests

are by far preferred on in-service networks. Among objective tests, the further distinction between intrusive (active) and non-intrusive (passive) solutions is introduced. When intrusive strategies are employed, test calls are deliberately injected in the network, to some extent spoiling its operating conditions; for passive solutions however, quality is inferred indirectly, from current network parameters such as packet loss rate, packet delay and jitter.

VQmon® [14]-[15] is the objective, non-intrusive tool employed in this study; it is an extension of the the E-Model [16], a well-established method for assessing the end-to-end transmission quality of a voice call. Exactly like for the E-Model, VQmon® output is a number between 0 and 100, the so-called Rating factor, R-factor for short, representing the overall call quality. The R-factor can be suitably mapped to the Mean Opinion Score (MOS)[17] on the well-known $1 - 5$ scale. Such correspondence has been recently updated for the Wide-band version of the E-Model [18], where the R-factor can reach values up to 129, as it might happen the AMR-WB codec is used. Table I summarizes the correspondence between the R-factor ranges and the MOS values, together with the classes of speech quality and user satisfaction.

VQmon® aims to seize the time-varying nature of packet losses, that heavily affect the quality of VoIP calls. For this reason, it extends the E-model, and rather than employing the average packet loss rate, it assumes there are two states of packet loss during the call: a high loss, burst state, and a low loss, gap state, each with a distinct packet loss probability.

III. SETTINGS AND RESULTS

A. Data Collection

We conducted this study on a pool of more than ten million VoLTE calls performed over a few days during the first half of 2018, in an urban area.

A single commercial LTE network was considered for the measurements. A proprietary probe was positioned at its Mb interface: the RTP voice flows traversing the interface were anonymized and inspected; the results were next aggregated in a *.csv* file. Positioning the tapping point at the Mb interface allowed to collect call detail records for both directions, i.e., for the voice flows being originated by the UEs and for the flows directed to the UEs, not necessarily originated within the same LTE network. For every call, we chose to analyze the uplink direction, in order to capture the negative effects that the Radio Access Network (RAN) traversal has on voice packets. For each call and for each direction, several data were available, such as the total number of transmitted packets, the

total number of received packets, the average and maximum jitter, the R-factor computed according to VQmon®, the type of codec, the call duration.

Moreover, a jitter buffer emulator (JBE) was instantiated, in order to realistically model the compensation that takes place receiver side, smoothing out the delay variations that voice packets exhibit after traversing the network. The emulator forces a delay on packets that arrive early, and immediately forwards late packets. In our system, the JBE was set to receive initial packets with a 50 ms delay, then to dynamically modify its play-out delay according to the average jitter of the previous 16 packets. Under these assumptions, we were able to estimate the packet loss rate, evaluating the ratio between the number of lost/excessively delayed packets and the total number of received packets after the JBE.

Filtering out invalid data and neglecting the calls that either employed the Enhanced Voice Services (EVS) wideband codec [19] or alternative, less popular speech audio codecs, we were left with $10,862,591$ voice calls. They were further distinguished in AMR and AMR-WB based, amounting to 71% and 29%, respectively. A reasonable explanation of the outstanding prevalence of AMR based calls lies in the inability of one party to support AMR-WB: in that event, AMR is chosen. It is on this number of calls that we based our analysis, first focusing on the LTE network conditions, as discussed next.

B. Network Conditions

To have an exhaustive picture of the operating conditions guaranteed to conversational voice by the cellular network where VoLTE calls were collected, we first computed the Cumulative Distribution Function (CDF) of the packet loss rate after the JBE and the CDF of the maximum jitter.

Fig. 2 shows the CDF of the packet loss rate P_{loss} experienced by the examined flows, for $0 \leq P_{loss} \leq 0.2$. This Figure indicates that P_{loss} values lower than or equal to 0.1 are guaranteed with probability 0.96 and that the probability an RTP stream experiences no packet losses is equal to 0.7, a remarkably high value, suggesting the LTE network under examination guarantees good operating conditions.

Last conclusion is corroborated by next figure. In detail, Fig. 3 reports the CDF of the maximum jitter J_{max}, for $0 \leq J_{max} \leq 1050$ ms, and shows that a stream experiences a maximum jitter value lower than or equal to 150 ms with a 0.9 probability.

Next, Figs. 4(a) and (b) provide a unified view of the examined LTE network: Fig. 4(a) portrays the joint probability density function (pdf) of the packet loss rate P_{loss} and of the maximum jitter J_{max} that the AMR based voice flows undergo, whereas Fig. 4(b) reports the joint pdf referring to the AMR-WB flows. Although slight numerical differences are present, it is immediate to notice that both pdfs exhibit a remarkable densification near the origin, a strong clue of good network functioning. Overall, these figures offer an enlightening spot on the QoS level that $QCI = 1$ services
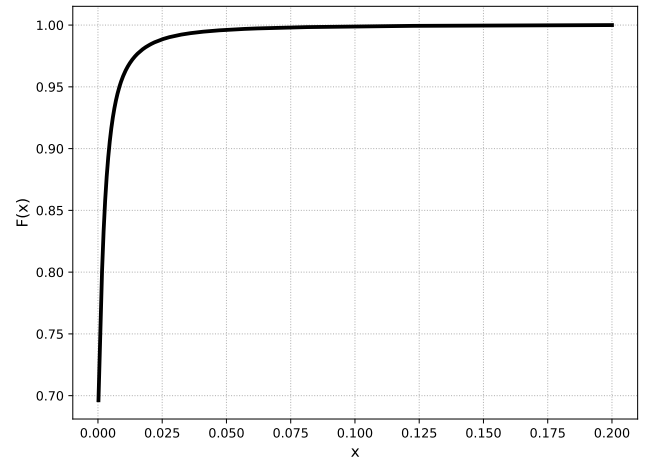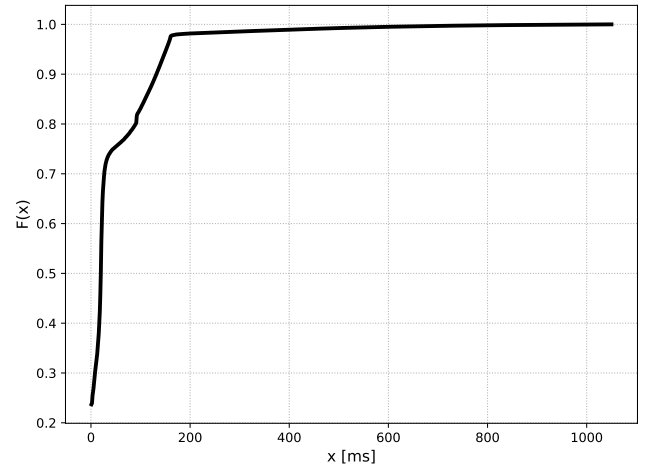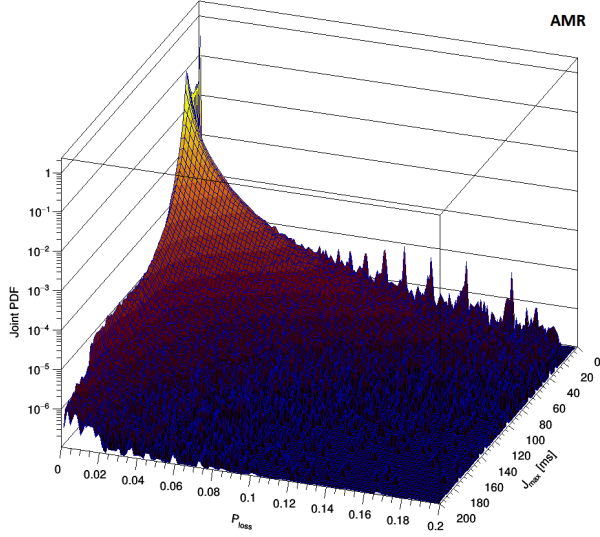

Fig. 2. Packet loss rate CDF

Fig. 3. CDF of the maximum jitter

experience in LTE, when clients are equipped with a playout buffer.

C. R-factor Results

Next goal was to investigate the QoE perceived by VoLTE calls. To this regard, Figs.5(a) and (b) display the R-factor values of the examined flows as a function of P_{loss} and J_{max}, for the AMR and AMR-WB case, respectively. The comparison between the two figures indicates that the adoption of the AMR-WB codec leads to higher R-factor values and suggests a far more pronounced dependence of the R-factor on P_{loss} than on J_{max}. Note that the jagged behavior appearing in Fig.5(b) is exclusively due to the lack of points in the region of high packet loss rates and high values of maximum jitter, that the former Figs.4 (a) and (b) already evidenced.

In order to better investigate the R-factor dependency on P_{loss}, Fig. 6 reports its average and standard deviation values over 10 uniform intervals of packet loss rate, when P_{loss} varies between 0 and 0.2. It is interesting to observe the sharp decay that the average R-factor displays for increasing P_{loss} values in the AMR case, whereas the decrease is less pronounced in the AMR-WB case. The standard deviation tends to increase for increasing values of the packet loss rate,
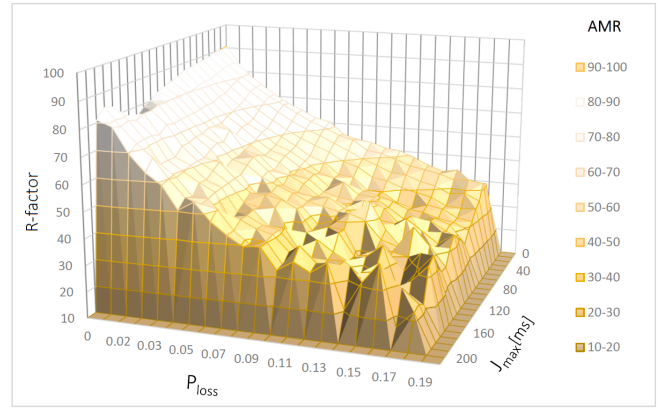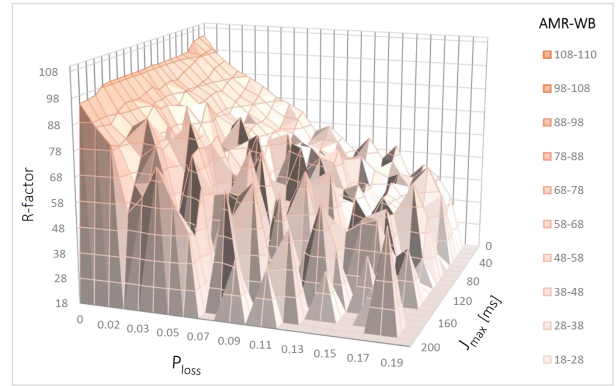
(a) AMR based flows
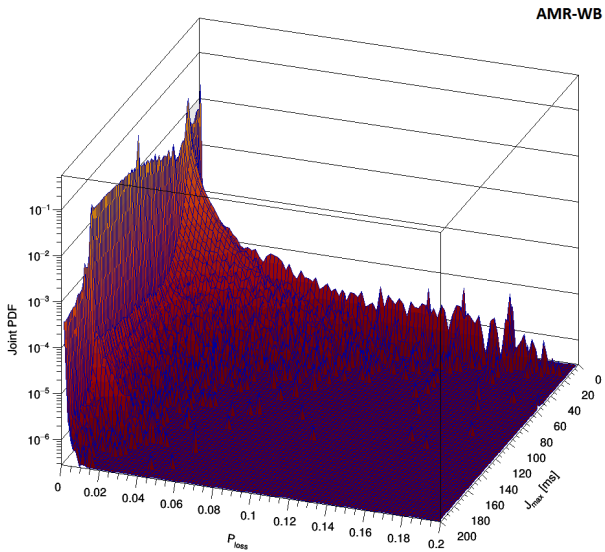
(b) AMR-WB based flows

Fig. 4. Joint pdf of the packet loss rate P_{loss} and of the maximum jitter J_{max}

(a) AMR based flows

(b) AMR-WB based flows

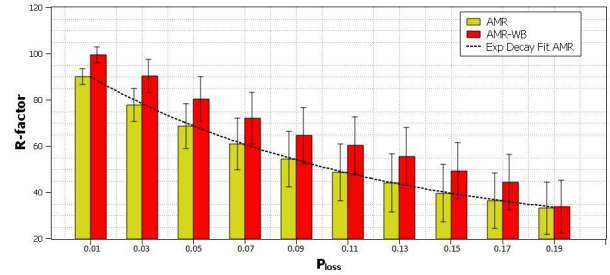Fig. 5. R-factor as a function of the packet loss rate P_{loss} and of the maximum jitter J_{max}

Fig. 6. R-factor as a function of P_{loss}

but this has to be mainly ascribed to a decreasing size of the population of samples. For the AMR case, this figure shows the first order, exponential fit performed on the set of (x_i, y_i) points, $i = 1, 2, \ldots, 10$, where x_i represents the median value of P_{loss} in every interval and y_i the value of the corresponding average R-factor. The Levenberg-Marquardt algorithm has been used, choosing $y(x) = y_0 + Ae^{-x/B}$ as the fit function (dashed line). The y_0, A and B values are 17.953, 71.63 and 0.12, respectively. By visual inspection, we

conclude that the fitting is truly satisfying. To confirm the goodness of the exponential choice, we computed the Mean Square Error (MSE), that measures the distance between the points estimated by the regression and the measured values: it turns out that $MSE_{exp} = 2.03$. Although not reported on the figure, we also tested the linear regression on the same set of data; the latter has been performed using $y(x) = A' + B'x$ as the fit function, with $A' = 85.74$ and $B' = -304.58$. For the linear regression $MSE_{lin} = 13.82$, confirming that the exponential fitting is by far better.

For the AMR-WB case, Fig.7 reports the original (P_{loss}, R) points, together with the comparison between the exponential and the linear fitting. For the former, $y_0 = -31.72$,

$A = 135.07$ and $B = 0.28$, whereas for the linear regression $A^{'} = 99.01$ and $B^{'} = -340.7$. Interestingly, in this case we have $MSE_{exp} = 4.25$ and $MSE_{lin} = 5.94$; moreover, for the linear case the coefficient of determination quantifying the fitting goodness is $R^2 = 0.98$. As a matter of fact, the linear choice is as adequate as the exponential. This can be explained observing that the R-factor decrease for increasing values of P_{loss} is much smoother when the AMR-WB codec is employed, than when AMR is.
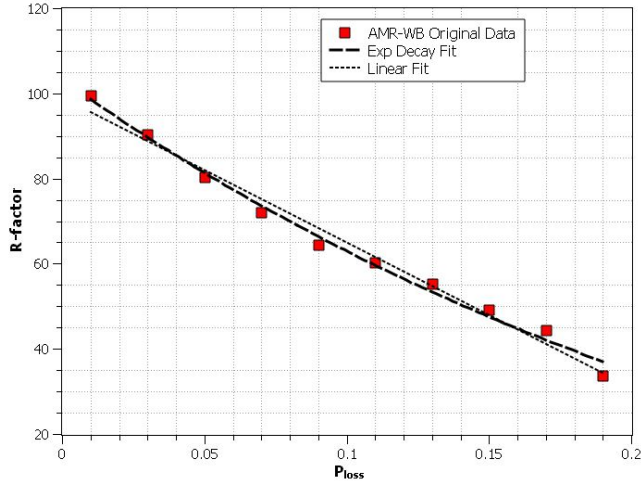

Fig. 7. Fitting comparison for the AMR-WB case

Although not reported in this paper, we verified that the dependency of the R-factor on J_{max} is pronounced for the AMR-based voice calls, whereas it is nearly absent for the AMR-WB based calls.

It is then possible to conclude that the quality experienced by the calls based on both codecs significantly depends on the packet loss rate values of the traversed LTE network. Furthermore, the R-factor dependency on P_{loss} can well be described by the exponential function for AMR based calls, whereas either a linear or an exponential decay captures such behavior for AMR-WB based calls. Overall, the above results reasonably allow to conclude that the quality dependence of AMR VoLTE calls on P_{loss} replicates the QoE – exponential – dependence on the QoS parameters first outlined in [3].

IV. CONCLUSIONS

This work has first portrayed the operating conditions that VoLTE calls experience on a real LTE commercial network, exploring the occurrence frequencies of the packet loss rate and maximum jitter values, i.e., of two amongst the most meaningful network parameters for real-time services. Next, with the help of an objective, no-reference metric, it has investigated the QoE guaranteed to AMR and AMR-WB based calls.

Examining over ten million calls, the study has revealed that the loss rate and the maximum jitter are successfully confined for VoLTE services and that the packet loss rate is the most relevant impairment to consider for both AMR and AMR-WB

calls. It has further demonstrated that the dependency of the R-factor on P_{loss} is successfully captured by an exponential law for the calls performed via the AMR codec, whereas either a linear or an exponential decay serve the purpose for the AMR-WB case.

REFERENCES

[1] 3GPP, "Circuit switched (CS) fallback in Evolved Packet System (EPS)," *3GPP TS 23.272, V. 15.0.0*, 2017.
[2] GSMA, "VoLTE Service Description and Implementation Guidelines," *GSMA Official Document FCM.01*, 2014.
[3] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, March 2010.
[4] J. Holub, M. Wallbaum, N. Smith, and H. Avetisyan, "Analysis of the Dependency of Call Duration on the Quality of VoIP Calls," *IEEE Wireless Communications Letters*, pp. 1–1, 2018.
[5] H. Taddei, I. Varga, L. Gros, C. Quinquis, J. Y. Monfort, F. Mertz, and T. Clevorn, "Evaluation of AMR-NB and AMR-WB in packet switched conversational communications," in *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, June 2004, pp. 2003–2006 Vol.3.
[6] Y. J. Jia, Q. A. Chen, Z. M. Mao, J. Hui, K. Sontineni, A. Yoon, S. Kwong, and K. Lau, "Performance characterization and call reliability diagnosis support for voice over LTE," in *MobiCom*. ACM, 2015, pp. 452–463.
[7] A. Elnashar, M. A. El-Saidny, and M. Mahmoud, "Practical Performance Analyses of Circuit-Switched Fallback and Voice Over LTE," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1748–1759, Feb 2017.
[8] J. Abichandani, J. Baenke, M. S. Irizarry, N. Saxena, P. Vyas, S. Prasad, S. Mada, and Y. Z. Tafesse, "A comparative study of voice quality and coverage for voice over long term evolution calls using different codec mode-sets," *IEEE Access*, vol. 5, pp. 10 315–10 322, 2017.
[9] 3GPP, "Policy and charging control architecture," *3GPP TS 23.203, V. 15.4.0*, 2018.
[10] ——, "Mandatory speech codec speech processing functions; AMR speech codec; General description," *3GPP TS 26.071, V.15.0.0*, 2018.
[11] K. Jarvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J. P. Adoul, "GSM enhanced full rate speech codec," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, Apr 1997, pp. 771–774 vol.2.
[12] 3GPP, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description," *3GPP TS 26.171 V.15.0.0*, 2018.
[13] ITU-T, "Ls on audio issues," *SG 16 Rapporteurs Q7/12*, 2006.
[14] A. Clark, "Modeling the effects of burst packet loss and recency on subjective voice quality," 2001.
[15] "Telchemy's VQmon," http://www.telchemy.com/vqmon.php, accessed: 2018-09-15.
[16] ITU-T, "The E-model: a computational model for use in transmission planning," *ITU-T Recommendation G.107*, 2015.
[17] ——, "Methods for subjective determination of transmission quality," *ITU-T Recommendation P.800*, 1996.
[18] ——, "Wideband E-model," *ITU-T Recommendation G.107.1*, 2015.
[19] 3GPP, "Codec for Enhanced Voice Services (EVS); General Overview," *3GPP TS 26.441, V. 15.0.0*, 2018.