DOCTORATE SCHOOL IN
INFORMATION AND COMMUNICATION TECHNOLOGIES

XXXII CYCLE

UNIVERSITY OF MODENA AND REGGIO EMILIA

ENGINEERING DEPARTMENT “ENZO FERRARI”

Ph.D. DISSERTATION

The role of Quality of Experience and Voice in 5G Networks

Candidate:
Elena CIPRESSI
Advisor:
Prof. Maria Luisa MERANI
School Director:
Prof. Sonia BERGAMASCHI

SCUOLA DI DOTTORATO IN INFORMATION AND COMMUNICATION TECHNOLOGIES

XXXII CICLO

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

DIPARTIMENTO DI INGEGNERIA “ENZO FERRARI”

TESI PER IL CONSEGUIMENTO DEL TITOLO DI DOTTORE DI RICERCA

Il ruolo della Quality of Experience e della Voce nelle reti 5G

Candidata:
Elena CIPRESSI
Relatore:
Prof. Maria Luisa MERANI
Il Direttore della Scuola:
Prof. Sonia BERGAMASCHI

"Io veggio ben che giá mai non si sazia nostro intelletto, se 'l ver non lo
illustra di fuor dal qual nessun vero si spazia.
Posasi in esso, come fera in lustra, tosto che giunto l'ha; e giugner puollo:
se non, ciascun disio sarebbe frustra.
Nasce per quello, a guisa di rampollo, a pié del vero il dubbio; ed é natura
ch'al sommo pinge noi di collo in collo."
Divina Commedia, Paradiso, Canto IV, vv. 124-132

Keywords:

5G
Voice over IP
Quality of Experience
Quality of Service
Network Measurements

Introduction

The advent of 5G networks is not only a technological revolution, it will also encompass new business models, whom service providers and network operators should be ready to adapt to. Differently from 4G, which enhanced the performance of the existing 3G systems, 5G will enable new communication paradigms such as device-to-device, inter- and intra- vehicular communications, as well as new services like remote healthcare and smart homes, just to name a few. 5G will dedicate different portions of the network to ad-hoc services with specific speed/latency requirements, taking advantage of the so-called "network slicing" approach that allows to deploy several multi-service, logically distinct networks over the same physical infrastructure. End-users and their expectations will be the focus of any provider. As such, the key asset of 5G will no longer be Quality of Service (QoS), rather, Quality of Experience (QoE). To anticipate market needs, the definition of a trustable monitoring approach to QoE is an action for Empirix, the company where I worked during my PhD years, as this company offers troubleshooting and diagnostics solutions to service providers and enterprises. This dissertation focuses on the exploration of 5G and QoE topics and in its final part, it proposes a novel technique to assess the quality that end-users will experience. Firstly, I followed 5G standardization with a special emphasis on investigating aspects related to Empirix core business. I transferred my knowledge to the whole company through trainings and lessons. After this phase of study, Empirix and I recognized the delivery of voice as one of the well-established and key feature that must be maintained in 5G applications, worthy of an in-depth study. Recently, Forbes stated: "*with 5G every object could soon have a voice*", and, according to Ericsson forecasting: "*voice is the king of communication and in a 5G world it will be more important than ever [...]. The network infrastructure used for Voice over LTE (VoLTE) today will also be used to enable 5G voice calls*". Thus, by exploiting passive network measurements, I performed a comparative study of the end-to-end quality degradation that several millions of real VoLTE calls underwent, when two popular codecs were employed, namely, Adaptive MultiRate (AMR) and Adaptive MultiRate WideBand (AMR-WB). This study revealed to what extent AMR-WB based calls are more robust against network impairments of voice services (e.g., jitter and packet loss rate) than their narrowband counterpart. In parallel, I also came to the conclusion that relying on empirical models to evaluate QoE leads to several limitations, most notably, the lack of actual feedback from end-users. I addressed this shortcoming adopting a customer-driven approach, although in a much more confined environment. To this regards, I leveraged a pool of research participants that listened and scored AMR-WB calls generated by Hammer, an Empirix platform that emulates Voice over IP communications, therefore providing truly subjective evaluation scores. By comparing different state-of-the-art algorithms my goal was to exploit statistical approaches based on supervised Machine Learning to understand how QoE was related to network metrics and human-based features like age, gender and type of headset. The major findings were: i) Ordinal Logit Regression is the algorithm that best captures the aforementioned relation. ii) Users usually agree to rate call quality as either "excellent" or "bad". iii) Conversely, when they are asked to rate call quality on the conventional five score scale, the statistical prediction of QoE becomes a difficult task: the perception of "intermediate" quality is deeply related to subjective personal traits and it may significantly vary from person to person.

Introduzione

L'imminente avvento del 5G non rappresenta solo una rivoluzione tecnologica, ma definisce anche nuovi modelli di business, a cui i service providers e gli operatori di rete dovranno adattarsi. A differenza del 4G, che ha migliorato in termini di performance il preesistente 3G, il 5G promette di abilitare nuovi paradigmi di comunicazione come le comunicazioni device-to-device, intra e inter-veicolari, cosí come di diffondere nuovi servizi, la tele-medicina e le case intelligenti, per citarne alcuni. Coadiuvata dal network slicing, che consentirá lo sviluppo di multiple reti logiche sulla stessa infrastruttura fisica, il 5G allocherá porzioni di rete a servizi ad-hoc con specifici requisiti di latenza e velocitá. L'utente sará al centro della rete, e con esso ogni sua aspettativa. Per tale motivo, il *key asset* del 5G non é piú la Quality of Service (QoS) ma la Quality of Experience (QoE). Per anticipare e rimanere competitivi nel mercato, la definizione di un sistema affidabile di monitoraggio della QoE é cruciale per l'azienda in cui ho svolto il mio PhD, Empirix, la quale offre servizi di diagnostica e troubleshooting ad operatori di rete ed imprese. Questa tesi esplora il tema del 5G e della QoE, e nella parte finale propone un approccio innovativo volto a determinare la qualitá percepita dall'utente finale. In una prima fase, ho seguito la standardizzazione del 5G, con particolare attenzione agli aspetti legati al core business dell'azienda. Ho svolto seminari e lezioni sulle principali caratteristiche del 5G. In seguito, assieme ad Empirix, ho individuato nei servizi voce, una delle applicazioni piú promettenti dell'ecosistema 5G, e meritevole di approfondimento. Recentemente, Forbes ha dichiarato che "*with 5G every object could soon have a voice*", e secondo le previsioni di Ericsson "*voice is the king of communication and in a 5G world it will be more important than ever [...]. The network infrastructure used for Voice over LTE (VoLTE) today will also be used to enable 5G voice calls*". Ho dapprima sfruttato misure passive di rete per analizzare e confrontare la qualitá di diversi milioni di chiamate VoLTE di una rete reale, codificate dagli odierni Adaptive Multi-Rate (AMR) e Adaptive Multi-rate Wide-Band (AMR-WB) codecs. Questa analisi ha permesso di verificare la maggiore robustezza delle chiamate AMR-WB rispetto agl'impariments dei servizi voce, quali jitter e packet loss rate. In parallelo, sono giunta alla conclusione che fare affidamento ad un modello empirico per monitorare la QoE porti a diverse limitazioni, prima fra tutte, la mancanza feedback dell'utente finale. Per ovviare a questa lacuna, ho adottato un approccio *customer-driven*, seppure in un contesto molto piú confinato del precedente. A questo proposito, ho sottoposto ad un gruppo di volontari un esperimento di valutazione di qualitá chiamate AMR-WB, generate in ambiente virtuale da Hammer, una piattaforma proprietaria di Empirix volta ad emulare comunicazioni Voice over IP. Confrontando le caratteristiche di diversi algoritmi dello stato dell'arte, ho allenato un modello di Machine Learning supervisionato per comprendere come la QoE fosse in relazione con le metriche di rete, e le caratteristiche dell'utente (etá, genere, tipo di dispositivo impiegato). Ho constatato che: i) l'algoritmo che meglio approssima la relazione di cui sopra l'Ordinal Logit Regression; ii) gli utenti spesso convengono nel definire una chiamata di qualitá "eccellente" o "pessima" iii) per contro, quando valutano una chiamata di qualitá definibile "intermedia", entrano in campo variabili soggettive che difficilmente possono essere interpretate in termini statistici, rendendo pertanto ardua la predizione di QoE.

Acknowledgments

My deepest gratitude is addressed to Prof. Maria Luisa Merani for her scientific mentoring and for accepting me as a PhD student. She gave me a solid scientific basis that I will take with me for the next steps of my career. A special thanks goes to Prof. Walter Cerroni (University of Bologna, Italy) and Prof. Ilenia Tinnirello (University of Palermo, Italy) for reviewing this dissertation and for sharing constructive suggestions to improve it.
I would like to thank all Empirix family for having welcomed me as a PhD apprentice, John d'Anna, Chief Executive Officer, Franco Messori, Chief Strategy Officer, and Vince Fortin, Senior VP of Engineering, for promoting my research activity within the company; I want express my sincere thanks to my mentor Santino Mele, for all the extra-time he dedicated to my activity and his guidance; my supervisor Andrea Malfitano for his technical training; Giovanni Tosatti for the product lessons; all the Engineering & Product Management mates and each colleague for their friendly support.
Outside from Empirix, my research activity has been encouraged by my pillars, lifelong friends and family. A special mention to the Trebbi family, for their close friendship during the last, though, year.
Last but not least, the biggest thanks goes to my "rocket scientist" Giovanni, who became my husband during this journey. He is the best part of me, believing in myself more than I usually do. This big achievement would not have been possible without him.

Contents

List of Figures

List of Tables

Chapter 1

An Introduction to 5G Networks

The goal of this chapter is to describe the key features of 5G Networks. Starting from the initiatives of 5G standardization bodies, and the time-line defined for 5G deployment, I will illustrate the demanding network requirements to support a brand-new set of use cases and applications. I will then investigate the technical aspects of 5G New Radio, with a special emphasis on the migration paths network operators might follow to roll-out their commercial networks. Next, I will summarize the architectural principles of 5G core, along with the new functionalities provided by Network Data Analytics Function (NWDAF) and the rising concept of Network Slicing. This chapter will end with an overview of the new emerging communication paradigms that will uniquely characterize the next generation of networks.

1.1 Actors for 5G Standards and 5G Standardization Process Involvement

Third Generation Partnership Project (3GPP) and International Telecommunication Union (ITU) are the main actors leading the complex process of 5G standardization. This section is intended to illustrate the relations among these entities, with different roles and responsibilities, and to give an overview of the reference standards that set requirements for the envisioned 5G use cases, the so-called International Mobile Telecommunications-2020 and beyond [1] (IMT-2020 for short).

1.1.1 3GPP and ITU

3GPP produces Technical Specifications, to be transposed by relevant Standardization Bodies (Organizational Partners) into appropriate deliverables (e.g., standards). Hence, 3GPP provides inputs to ITU, that releases the standards. The standards for 5G will be under the IMT-2020 umbrella. The technologies that are currently of interest for 3GPP are:

- Long Term Evolution-Advanced (LTE-A)
- LTE
- Carrier Aggregation Explained
- HetNet/Small Cells
- Non Access Stratum (NAS)
- Evolved Packet Core (EPC)
- High-Speed Packet Access (HSPA)
- Universal Mobile Telecommunication System (UMTS)
- Wideband-Code Division Multiple Access (W-CDMA)
- General Packet Radio Service (GPRS) & Enhanced GPRS (EDGE)

3GPP has seven Organizational Partners - from Asia, Europe and North America that determine the general policy and strategy of 3GPP. They are the following:

1. ARIB, The Association of Radio Industries and Businesses, Japan
2. ATIS, The Alliance for Telecommunications Industry Solutions, USA
3. CCSA, China Communications Standards Association
4. ETSI, The European Telecommunications Standards Institute
5. TSDSI, Telecommunications Standards Development Society, India
6. TTA, Telecommunications Technology Association, Korea
7. TTC, Telecommunication Technology Committee, Japan

3GPP adopts a two-phase based approach, comprising (i) Study Item(s) and (ii) Work Item(s). Phase 2 Study Item(s) will coincide with Phase 1 Work Item(s) beginning. Phase 1 corresponds to Release 15, that has been completed in September 2018. Within Phase 1, initial 5G deployments has been defined to address crucial subset of the commercial needs. Phase 2 corresponds to Release 16, that should be completed by March 2020 for IMT-2020 submission. Phase 2 will cover all the identified use-cases and requirements for 5G effective deployment actions.

The **ITU** with its "Working Party 5D (WP 5D) - IMT Systems" works on IMT-2020[1], the standard for 5G mobile systems. In this document "*Plan, Timeline, Process and Deliverables for the future development of International Mobile Telecommunications (IMT)*" are addressed. The objective of this Recommendation from the outset was to establish the vision for IMT for 2020 and beyond, by describing potential user and application trends, growth in traffic, technological trends and spectrum implications, and by providing guidelines on the framework and the capabilities for IMT for 2020 and beyond. In particular, WP 5D provides a detailed time-line for the standardization process[1]: "*in the 2016-2017 time-frame, WP 5D will define in detail the performance requirements, evaluation criteria and methodology for the assessment of new IMT radio interface. It is anticipated that the time-frame for proposals will be focused in 2018. In 2018-2020 the evaluation by independent external evaluation groups and definition of the new radio interfaces to be included in IMT-2020 will take place. Working Party 5D also plans to hold a workshop in late 2017 that will allow for an explanation and discussion on performance requirements and evaluation criteria and methodology for candidate technologies for IMT-2020 that has been developed by WP 5D, as well as to provide an opportunity for presentations by potential proponents for IMT-2020 in an informal setting.*" The whole process is planned to be completed in 2020 when a draft new ITU-R Recommendation with detailed specifications for the new radio interfaces will be submitted for approval within ITU-R.

1.1.2 The IMT-2020 Standard

The objective of IMT-2020 is supporting and expand various usage scenarios and applications that will continue beyond the current IMT. Moreover, a wide variety of key requirements would be tightly coupled with these different

[1]`https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx`

usage scenarios and applications for IMT-2020. Figure 1.1 illustrates some examples of identified usage scenarios for IMT-2020.

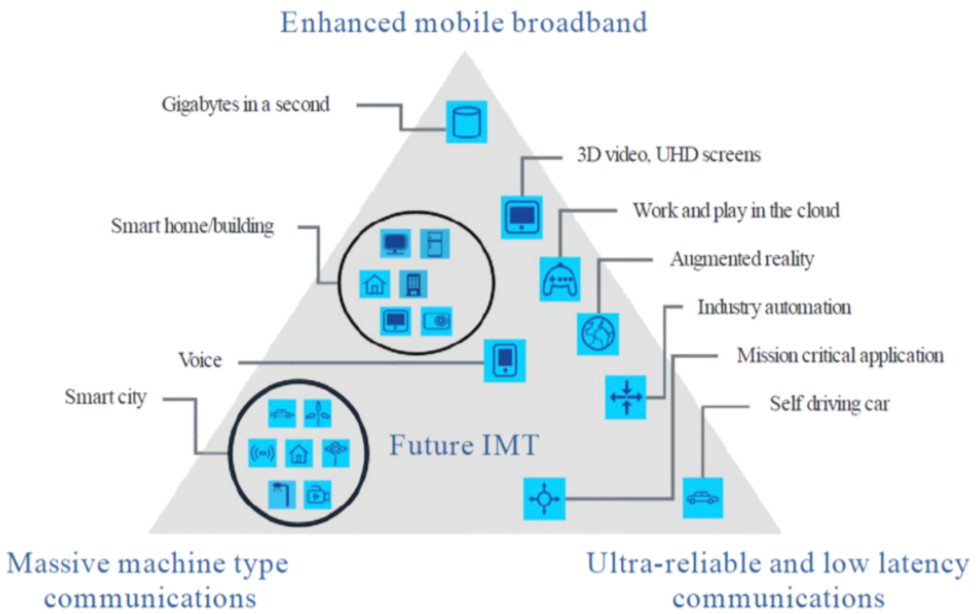


Figure 1.1: Usage Scenario for IMT 2020[1]

Taking Figure 1.1 as a reference, threes are the major usage scenarios, that include enhanced Mobile BroadBand (eMBB), Ultra-reliable and Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC). Many applications are in-between these three application areas.

- The eMBB applications are centered on the human needs to get access to multimedia content, services and data.
- URLLC applications have stringent requirements for capabilities such as latency and availability, thus supporting mission critical services such like autonomous vehicles and drone applications.
- mMTC applications represent the massive deployment of IoT applications, characterized by a large number of connected devices, typically transmitting relatively low volumes of low priority data. Enabled by low-cost, long life modules with sensors and connectivity, massive IoT applications will range from asset tracking, smart cities, monitoring of utilities and vital infrastructure.

1.2 5G Key Capabilities at a Glance

The envisaged usage scenarios and applications require a robust and flexible network infrastructure that could support such a heterogeneous ecosystem of applications. 5G key capabilities are summarized in Figure 1.2, where the enhancements with respect to LTE (in the ITU jargon, "IMT-advanced") are underlined.

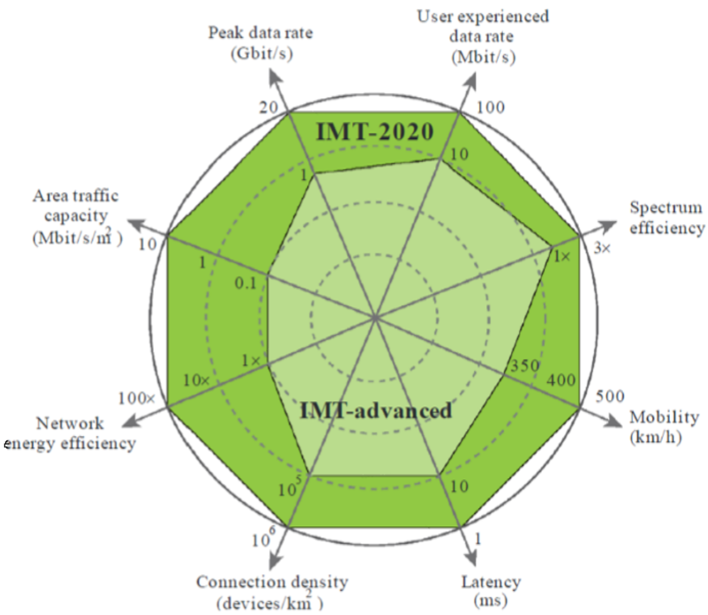


Figure 1.2: Enhancement of key capabilities from IMT-Advanced to IMT-2020[1]

In the vision of IMT-2020, eight are the parameters characterizing the capabilities of 5G Networks (Figure 1.2):

1. **Peak data rate**, defined as the maximum achievable data rate under ideal conditions per user/device (in $Gbit/s$). The peak data rate of IMT-2020 for eMBB is expected to reach 10 $Gbit/s$; for wide area coverage cases (e.g., in urban and sub-urban areas, $100Mbit/s$); $1Gbit/s$ in hotspot cases (e.g., indoor).

2. **User experienced data rate**, defined as the achievable data rate that is available ubiquitously across the coverage area to a mobile user/device (in $Mbit/s$ or $Gbit/s$). The achievable increase in efficiency from IMT-Advanced will vary between scenarios and could be higher in some scenarios (for example five times subject to further research).

IMT-2020 is expected to support $10Mbit/s/m^2$ area traffic capacity, for example in hot spots.

3. **Latency**, defined as the contribution by the radio network to the time when the source sends a packet to the time when the destination receives it (in ms). IMT-2020 will be able to provide $1ms$ over-the-air latency, capable of supporting services with very low latency requirements.

4. **Mobility**, defined as the maximum speed at which a seamless transfer between radio nodes can be achieved (in km/h). IMT-2020 is expected to enable high mobility up to $500km/h$ with acceptable quality. This is envisioned in particular for high speed trains.

5. **Connection density**, defined as the total number of connected and/or accessible devices per unit area (per km^2). IMT-2020 is expected to support a connection density of up to $10^6/km^2$, for example in massive machine type communication scenarios.

6. **Energy efficiency**, whose definition is two-fold: (i) on the network side, energy efficiency refers to the quantity of information bits transmitted to/ received from users, per unit of energy consumption of the radio access network (RAN) (in $bit/Joule$); (ii) on the device side, energy efficiency refers to quantity of information bits per unit of energy consumption of the communication module (in $bit/Joule$). The energy consumption for the radio access network of IMT-2020 should not be greater than IMT networks deployed today, while delivering the enhanced capabilities. The network energy efficiency should therefore be improved by a factor at least as great as the envisaged traffic capacity increase of IMT-2020 relative to IMT-Advanced for eMBB.

7. **Spectrum efficiency**, defined as the average data throughput per unit of spectrum resource and per cell ($bit/s/Hz$). The spectrum efficiency is expected to be three times higher compared to IMT-Advanced for eMBB. The achievable increase in efficiency from IMT-Advanced will vary between scenarios and could be higher in some scenarios (for example five times subject to further research). The minimum requirements for peak spectral efficiency are expected to reach $30bit/s/Hz$ (downlink) and $15bit/s/Hz$ (uplink).

8. **Area traffic capacity**, defined as the total traffic throughput served per geographic area (in $Mbit/s/m^2$). IMT-2020 is expected to support a area traffic capacity up to $10Mbit/s/m^2$.

1.3 5G Radio Access Network

This section aims to provide an overview of the technological aspects that 5G New Radio features, along with LTE radio device enhancements to meet New Radio expectations. The high demanding requirements described in section 1.2 need to be supported by a more sophisticated radio equipment that could operate in every portion of the spectrum, from the lowest frequencies (below $1GHz$) to the highest (above $6GHz$).

1.3.1 5G New Radio

The new 5G Radio Access Technology (RAT) is referred to as New Radio (NR).

There are a number of features that are unique for 5G radio access compared to the previous generations, such as a wide range of carrier frequencies and deployment options, diverse use cases with very different user requirements, small-size base stations, self-backhaul, massive multiple-input multiple-output (mMIMO), and large channel bandwidths. NR is intended to be optimized for performance without considering backward compatibility in the sense that LTE user equipments (UEs) do not need to be able to camp on an NR carrier.

As well described and summarized by A. A. Zaidi et al in [4], these requirements ask for a flexible waveform, numerology, and frame structure. NR has to support applications with very low latency, which needs very short subframes. NR should support both access and backhaul links by dynamically sharing the spectrum, enabling the full potential of multi-antenna technology. The number of antenna elements may vary, from a relatively small number of antenna elements in LTE-like deployments to many hundreds in NR, where a large number of active or individually steerable antenna elements are used for beamforming, single-user (SU-MIMO) and multi-user MIMO (MU-MIMO). 5G NR will feature a novel TDD/FDD design to deliver latency in the ms range and it will maintain backward compatibility with LTE Radio. NR is envisioned to mainly be based on time-division duplex (TDD) at high frequencies (above $6GHz$) and mainly on frequency-division duplex (FDD) at lower frequencies. At very high frequencies, base stations can be small (low-cost) access nodes, putting similar requirements in donwlink (DL) as in uplink (UL) (transmit power, hardware impairments, etc.). In March 2016, 3GPP agreed to study various features of NR assuming orthogonal frequency-division multiplexing (OFDM), currently used in LTE for DL transmission.

5G will make the wider usage of the spectrum than ever, enabling different applications by exploiting both low/mid/high bands and either portions of licensed/shared/unlicensed spectrum. Such a wide exploitation of the spectrum will open new debates on spectrum allocation on a global scale. Here, I provide a brief classification of applications by spectrum range:

- **Low-bands** (below $1GHz$): for IoT services and extension of mobile broadband coverage (suburban and rural areas).
- **Mid-bands** ($1GHz$ to $6GHz$): a reasonable mixture of coverage and capacity for 5G services, that will be used for initial 5G deployments.
- **High-bands** (above $6GHz$): mmWave spectrum. Even with poor penetration (in the order of mm), a very large bandwidth can be allocated to mobile communications, thus enabling enhanced mobile broadband applications.

1.3.2 LTE Radio Evolution towards 5G

LTE is also expected to evolve to capture a part of the 5G requirements. A tight integration of NR and LTE is envisioned in order to efficiently aggregate NR and LTE traffic. In view of this, 3GPP Release 13 has introduced some enhancements to LTE radio in order to ease the migration to 5G NR salient technological enhancements, that are summarized below:

- Active Antenna Systems (AAS) and associated Self Organized Network techniques. AAS systems can be considered as the basis for the so-called Full Dimensional MIMO (FD-MIMO) systems. An active antenna (AA) is a MIMO antenna that has active electronic components. The beam of an AA, driven by software, can be adjusted according to the capacity and coverage targets of the network. Beamforming is an AAS feature that allows the network service to adapt to changing situations in the cellular network, providing a more efficient way to serve parallel users.
- Elevation Beamforming and Full-Dimensional MIMO (FD-MIMO). The MIMO enhancements in 3GPP makes possible to smartly adapt transmission both vertically and horizontally by utilizing a steerable two-dimensional antenna array. FD-MIMO simultaneously supports 3D (elevation & azimuth) beamforming and more than 10 UEs Multi User (MU)-MIMO.
- Coordinated Multi-Point transmission and reception (CoMP). CoMP refers to a wide range of different techniques with the common trait

of dynamic coordination of transmission and/or reception at multiple geographically separated sites. Thus, in a distributed approach environment with no central coordinating nodes, CoMP is addressed to exchange control information among nodes in a coordinated set. The purpose of the coordination among cells is mitigating and then exploiting inter-cell interference, in order to enhance system performance and end-user service quality.

- Licensed Assisted Access (LAA) using LTE. This feature leverages the 5GHz unlicensed band in combination with licensed spectrum, to deliver a performance boost for mobile device users. As a matter of fact, 5G will move upward relying on LTE cross-carrier control mechanism.

1.3.3 Target Deployment Scenarios

Differently from previous generations, that demanded for both access and core network of the same generation, with 5G it is possible to integrate elements of different generations in different configuration, namely Non-Standalone and Standalone scenarios, whose salient features are outlined below:

- **Non-Standalone (NSA)**, features multiple radio access technologies to provide radio access: the NR radio cells are combined with LTE radio cells using dual connectivity. Depending on the choice of operator, the core network may be either EPC or 5G Core. This scenario requires tight interworking with the LTE RAN.

- In **Standalone (SA)**, only one radio access technology is used: NR/evolved LTE radio cells are exploited for both Control Plane and User Plane. This option may be deployed as an independent network using normal inter-generation handover between LTE and 5G for service continuity. SA will open the door to greenfield operators that might want to deploy the network from scratch.

Overview of NSA and SA Options

On December 2017, within Release 15, the first 5G NR specifications has been approved. Among Non-Standalone / Standalone configurations, 3GPP has identified a set of target deployment scenarios that network operators should support in the short, medium-long term to ease the migration to 5G networks.

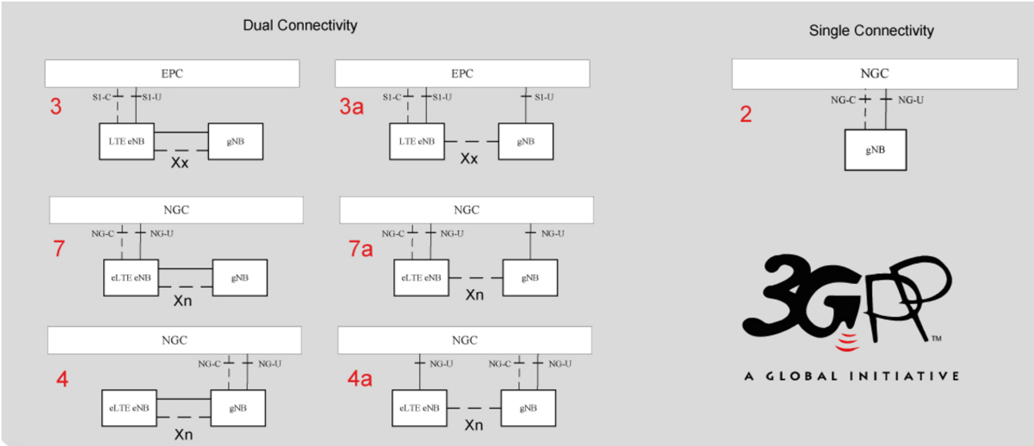


Figure 1.3: Taxonomy of NSA Options [2].

Taking Figure 1.3 as a reference, dotted lines represent Control Plane (CP) signaling, while continuous lines represent User Plane (UP) connections. New Radio terminology has been introduced in the 3GPP working group RAN3-NR2[2], and here a brief summary of this new lexicon is provided:

- NGC: Next Generation Core, namely the 5G Core (5GC)
- NG-RAN: a Radio Network which supports either NR or E-UTRA or both, interfacing with Next Generation Core (NGC). The NG-RAN consists of a set of gNBs connected to the 5GC through the NG
- NR gNB: equivalent of LTE eNB in 5G NR. A gNB can support FDD mode, TDD mode or dual mode operation
- eLTE eNB: evolution of eNB that supports connectivity to EPC and NGC

In this classification, the concepts of Master Node (MN) and Secondary Node (SN) are fundamental. MN establishes a direct CP connection with the Core (either be EPC or 5GC), while SN establishes a CP connection with the Core Network only through MN.

- Option 3: Non-standalone LTE and NR under EPC. The E-UTRAN is extended to allow compatible devices to use Dual Connectivity to combine LTE and NR radio access. Standard term for Option 3 is EN-DC, that stands for "E-UTRAN New Radio-Dual Connectivity" [5].

[2]Revision of R2-172641

One of the key advantages of this option is that it only requires the development of specifications of NR as non-standalone access as part of E-UTRAN connected to EPC rather than the specification of the full 5G system.

- Option 2: Standalone NR under 5GC. SA Option 2 envisages the deployment of both NR gNB based NG-RAN as a new radio access and 5GC as new core along with new features on LTE eNB based E-UTRAN to support inter-RAT mobility. Option 2 requires the device to support both a radio front end capable of receiving and transmitting data over NR as well as new procedures for the 5GC. The UE supports complete set of functionalities for CP and UP and for all interfaces to the network.

- Option 7: Non-standalone LTE and NR under 5GC. LTE RAN needs upgrade to connect to 5GC and more LTE base stations (eNode B) may need to be upgraded to interwork with NR. This option allows operators to continue to selectively deploy NR only where needed. As LTE is already offered in wide-area coverage in initial condition (NSA Option 3), the network can still leverage the wide-area coverage LTE network and deploy NR only when intended use case requires it.

- Option 4: Non-standalone NR and LTE under 5GC. LTE RAN needs upgrade to connect to 5GC and more LTE base stations (eNodeB) may need to be upgraded to interwork with NR. This option allows operators to continue to selectively deploy NR only where needed. However, compared with Option 7, this path may require the deployment of a larger number of more NR gNB since NR acts a MN, with a LTE SN, in the area where Option 4 is to be used.

Short-term Possible Migration Paths

Adopting the view of GSMA [6], a network operator may follow alternative migration paths towards the commercial 5G deployment, for the three-years view of 2019/2021.

- From LTE to NSA Option 3: Besides the accelerated time to market, as the NR will augment the existing capability of the LTE radio network, this option allows flexible "on demand" deployment where capacity is needed using the same or different vendors for LTE and NR. Furthermore, this option is going to be maintained in future releases of 3GPP (beyond Release 15) and therefore can be used in longer-term, even if other options are deployed in parallel. The capability of deploying NR while anchoring the communication to the EPC network offers the opportunity of making optimal use of the spectrum above $6GHz$ where operators will have available the large bandwidths necessary to deliver the high throughput in hotspots but that cannot be provided easily over large areas due to the fast signal attenuation.

- From NSA Option 3 to SA Option 2. The operator might migrate from having only NSA Option 3 to adding SA Option 2 with inter-RAT mobility mechanism used to move devices between 5G NSA LTE plus NR under EPC coverage, and 5G NR under 5GC coverage. Whereas the network was not able to leverage the advantages of 5GC in NSA Option 3, in this scenario the full advantage of 5G end-to-end network capabilities can be delivered to the users. This path enables network operators to address all use cases envisioned by ITU in IMT-2020.

- From LTE to SA Option 2. The operator might migrate directly to the SA Option 2 with inter-RAT mobility mechanisms used to move devices between 4G LTE under EPC coverage and 5G NR under 5GC coverage. One of the key benefit of this option is that SA architecture can take full advantage of 5G end-to-end network capabilities supported by NR and 5GC, providing customized service, especially to vertical industry, in an effective and efficient way. New features, including service-based architecture, and end-to-end Network Slicing (that will be later discussed), can be enabled according to specific requirement of each service, thus providing a customized and superior user experience.

1.4 5G System Architecture

The 5GC should be flexible enough to support the heterogeneous use cases described before. Recent advances in mobile cloud computing infrastructure allowed scalable, on-demand access to a vast pool of configurable resources like processing speed, storage, networking and integrated applications over the Internet. This centralized operational model reduces cost, increases availability, disconnects services from the existing technology and offers flexibility in terms of provisioning. This is a major transformation for operators as they will change how they operate and deliver services. A cloud-native architecture allows the software to evolve independently from the hardware. This reflects in moving some of the key resiliency and failover mechanism to the cloud, thus supporting a flexible information models based on micro-services. Differently from previous generations, in such a software-oriented framework, the modularized functions can be invoked using a standard API. Three are the main architectural enablers and components that will be crucial in transforming the networks to a completely cloud-based architecture:

- *Cloud Radio Access Network (C-RAN)*, a novel mobile network architecture designed to address the challenges operators face while trying to support growing end-user numbers.

- *Software Defined Networking (SDN)* is an architectural framework for creating intelligent, flexible programmable networks by decoupling control and user forwarding functions.

- *Network Function Virtualization (NFV)* renders network functions once tied to specific hardware appliances to run on industry-standard infrastructure operating in any data center.

Main 5GC specification has been deferred until Phase 2 of Release 16, and they are currently under definitions. However, 3GPP addressed to LTE some of the key enhancements to facilitate network operators to migrate to 5GC, like Control and User Plane Separation (CUPS) of the EPC [7]-[8]. CUPS enables flexible network deployment and operation, by the independent scaling between CP/UP. CUPS allows for reducing latency on application services, as well as handling the increase of data traffic and it brings application hosting from centralized data centers down to network edge, closer to consumer and data generated by applications.

Since this shift in paradigm from traditional telecom infrastructure to IT service is crucial, and it could be cumbersome to implement, 3GPP has defined two different representations of the 5G Core itself, the so-called "Point-to-point Architecture" (telecom-oriented) and the "Service Based Architecture" (IT-oriented), both detailed in the 3GPP System Architecture reference [9], and outlined in the next section.

1.4.1 Reference-Point Representation

This architecture diagram shows the interactions that exist between Network Functions (NF). It is the well-established "3GPP-like" representation that helps to inspect the differences and the similarities with the EPC.

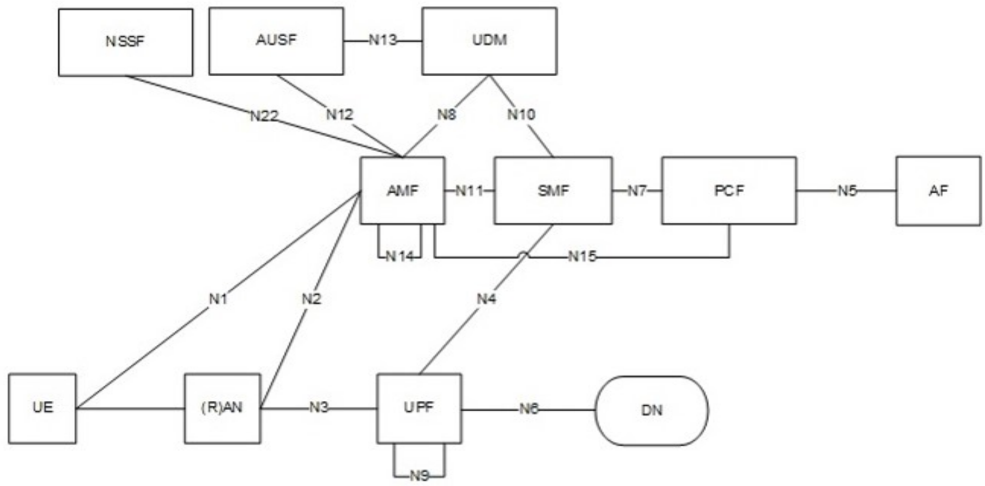


Figure 1.4: Non-Roaming 5G System Architecture in reference point representation [9]

The main 5G Core NFs are the following:

- Access Management Function (AMF). It manages the session control and the mobility (part of the MME functionalities).
- Session Management Function (SMF). It sets up and manages sessions according to network policy (part of the MME/SGW functionalities).
- User Plane Function (UPF). It can be deployed in various configurations and locations according to the service type (part of the SGW/PGW functionalities).
- Policy Control Function (PCF). It provides a policy framework for incorporating network slicing, roaming and mobility management (part of the PCRF functionalities).
- Unified Data Management (UDM). It supports the access authorization and subscription management (part of the HSS functionalities).
- Authentication Server Function (AUSF). It acts as an authentication server (part of the HSS functionalities).
- Network Slice Selection Function (NSSF). UE performs network assisted slice selection based on policy, via AMF. NSSF has not a LTE counterpart: I discuss later the concept of Network Slicing.

This representation shows the great complexity of adding new network elements/interfaces. The operators often require to reconfigure multiple adjacent interfaces. Thus, over the time, Reference-Point representation will be substituted by a Service Based model, using reusable APIs between any CP/UP functions. Existing mobile networks cannot customize control functions for a specific service type, like SBA does by ensuring short time to market for new services and greater flexibility for system updates.

1.4.2 Service Based Architecture

This representation provides a set of logical control function for diversified services. It illustrates how network functions within the CP enables other authorized NF to access their services. Service-based interfaces are used within the CP, with the purpose of simplifying the development and deployment of new services.

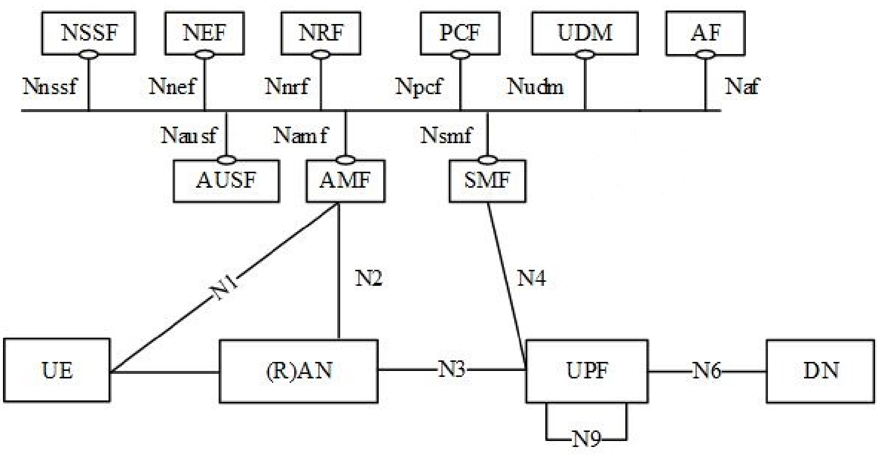


Figure 1.5: 5G System Architecture [9]

In this representation, additional NFs appear:

- Network Exposure Function (NEF). It provides a mean to securely expose the services and the capabilities provided by 3GPP NFs. Examples would include third party internal exposure or re-exposure.

- Network Repository Function (NRF). It supports the service discovery function. As such, it is able to receive NF discovery request from a NF instance and it can provide information about discovered NF instances.

1.4.3 The Concept of Network Slicing

Network Slicing technology is expected to be one of the game changer of 5G Networks. A network slice is a logical network serving a defined business purposes or customer, consisted of all required network resources configured together. A network slice is created, changed and removed by UE via AMF. As per 3GPP definition: [10] "*Network slicing allows the operator to provide customized networks. For example, there can be different requirements on functionality (e.g., priority, charging, policy control, security, and mobility), differences in performance requirements (e.g., latency, mobility, availability, reliability and data rates), or they can serve only specific users (e.g., Multimedia Priority Service users, Public Safety users, corporate customers, roamers, or hosting an Mobile Virtual Network Operator). A network slice can provide the functionality of a complete network, including radio access network functions and core network functions (e.g., potentially from different vendors). One network can support one or several network slices.*"
Network slice is not a new concept in the 3GPP specifications: in Release 13, the "DECOR" allowed the UE to connect to a Dedicated Core Network (DCN), based upon the parameter "*UE usage type*". In Release 14, under the evolved DECOR (eDECOR) framework, the UE assisted RAN in Access Stratum signaling towards the eNB during the EPC registration. However, both DECOR and eDECOR showed a limit: the UE can connect to only one DCN, that could not be optimized to multiple services. As a matter of fact, the slicing technology allows the operators to deploy multiple and independent end-to-end networks over the same network infrastructure.

1.4.4 The Network Data Analytics Function

There is another important function in 5GC, the Network Data Analytics Function (NWDAF) whose connections with other NFs, e.g. PCF, are not depicted in the reference point and service-based architecture diagrams. The NWDAF [3] has been introduced in the 5G System Architecture in 3GPP# SA2119 meeting, on February 2017. It is expected that will play a crucial role in NFs provisioning and management. As a matter of fact, NWDAF is responsible for transferring network analysis information when requested from any Network Functions, as depicted in Figure 1.6.

Analytics information are either statistical information of the past events, or predictive information. Different NWDAF instances may be present in the 5GC, with possible specializations per categories of analytics, and its capabilities are described in the NWDAF profile stored in the NRF. For instance, NWDAF could provide network slice level data analytics (e.g., load level in-

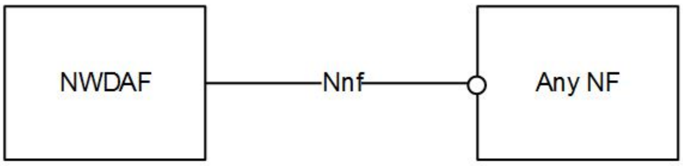


Figure 1.6: Data Collection architecture from any NF [3]

formation) to PCF and NSSF. PCF uses that data in its policy decisions, and NSSF may use the load level information provided by NWDAF for slice selection.

1.5 5G Emerging Communications

5G will open to new communications paradigms, that can be broadly classified in two groups: the Machine-to-Machine Communications and its extension, the Vehicle-to-Anything Communications.

1.5.1 Machine-to-Machine Communications

5G should ensure direct machine-to-machine communication (M2M for short) without infrastructures such as Access Points or Base Stations. This will be possible by optimizing the resource control policies, with more degrees of freedom in terms of mode selection, power control, and resource allocation. M2M communication will exploit portion of the "high-bands" spectrum, communications between devices will be ran in close vicinity (mmWave wavelength).

1.5.2 Vehicle-to-Anything Communications

It is expected that 5G will be used to enable all forms of extra vehicle communication, initially to provide more sophisticated advanced driver assistance systems and eventually leading to fully autonomous self-driving vehicles. In particular:

- V2I - Vehicle-to-Infrastructure: e.g. traffic signal, timing/priority
- V2N - Vehicle-to-Network: e.g. real time traffic/routing, cloud services
- V2P - Vehicle-to-Pedestrian: e.g. safety alerts to pedestrian, bicyclists
- V2V Vehicle-to-Vehicle: e.g. collision avoidance, safety systems

Using 5G technology to develop such a rich ecosystem scenarios will help to reduce costs and investment into infrastructure, which would be required by alternative technologies "dedicated" to just automotive applications.

Chapter 2

Quality of Experience towards 5G Networks

After the description of 5G architectural concepts and framework, in this chapter I am going to introduce the definition of Quality of Experience (QoE) and of its technological counterpart, the so-called Quality of Service (QoS). Then, I will underline the elements that are natively designed in 5G Networks to address and satisfy the user experience needs. Finally, I will shed light on the technological and architectural challenges 5G will have to deal with, in order to meet such high users' demands and expectations.

2.1 Quality of Experience and Quality of Service

There are different definitions of QoE across ITU, ETSI and others organizations' deliverables. For instance, ETSI [11] defines QoE as:

"A measure of user performance based on both objective and subjective psychological measures of using an Information and Communications Technology (ICT) service or product."

Whereas, in the Recommendation P.10/G.100 [12] ITU adopted the Qualinet definition [13]:

"The degree of delight or annoyance of the user of an application or service"

The ITU Recommendation [12] specifies in a note that *"Recognizing ongoing research on this topic, this is a working definition which is expected* *to evolve for some time"*, thus giving an idea of the complex research fields underneath this topic.

The technical report [11] enlightens how, in the last two decades, the use of ICT had extended from the workplace to the home and for applications that support leisure and social activities in addition to work. Consequently, the concerns of human-computer interaction have evolved from a focus on effectiveness and efficiency to user experience factors such as entertainment, engagement and the appeal of using and owning ICT. Whereas most of the work on user experience is conducted in relation to computer applications, there is also the need to address the user-centered development of telecommunication services. Telecommunication services are similar to computer applications since they require users to interact with devices and applications, with hardware and software interfaces. In addition, however, with telecommunication services, users have the specific intention to communicate with other people at distance. This communication is either direct to other people through technology or involves interaction with a machine rather than with a person. During either person-to-person (two-way) communication or person-to-machine (one-way) communication, the users will interact with a service that will have properties that may vary and that may have an effect on users behavior. For example, a delay between the arrival of audio and video information may lead to lack of lip-synchrony of the speaker as perceived by a listener. Properties such as audio-video asynchrony, transmission delay, video frame-rate and resolution have the potential to help or hinder communication. These technical properties are under the scope of the so-called Quality of Service (QoS).

Among service providers, network operators and equipment manufacturers, QoS has been in use for a long time and it has reached a high level of common understanding. QoS has its strongest reference from the ITU [14], where it has been defined as:

"Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service."

Although ITU refers to user satisfaction, the definition of QoS is well established and it is often used to determine the technical parameters of telecommunication applications such as network delay and packet loss. In addition to that, the focus on user satisfaction is rather limited because it is only one of many measures of user behavior with a communication service. For example, other measures include the time taken to perform a communication task (a measure of efficiency) and the accuracy with which a task is completed (a measure of effectiveness).

2.2 Benefits of QoE Data

Whereas the definition of QoS covers technical performance (i.e., it is mainly technology-centered), QoE is based on end-user behavior (i.e., it is user-centered). Working on QoS is critical, but not sufficient, for measuring user experience: QoE and QoS are distinct and both are important and should be interrelated. To this regard, ETSI Technical Report [11] identifies some essential factors related to the importance of collecting QoE-data:

- *To prevent churn.* Although developing a product or service that has instant appeal may increase the probability of a purchase and use, the profitability and image of the supplier can be predicted to be affected if subsequently the product or service does not meet up to expectation.

- *To prevent product or service rejections.* There is a history of products and services that have been rejected from the market despite that marketing departments have predicted success and without conclusively being able to explain the rejection. One important reason for rejection is that the QoE has been too low in the usage situation. Some of these rejections would have be foreseen, understood and avoided if QoE data had been applied or user tests been performed before product launches.

- *To optimize a product or service.* Within technical teams working with products, there may be little knowledge of how a certain set of technical parameters will be experienced by the end users. A particularly important situation is when trade-off decisions are required, such as with packet loss versus delay for speech services, frame-rate versus resolution for audio-visual services and bitrate versus latency for multimedia broadband services.

- *By expressing QoE as a function of QoS.* It is argued that the focus of QoE should be extended to include how end-users experience the use of a specific service, terminal or network. It is also argued that QoE data should succeed where possible, to combine knowledge of both user experience and technical parameter values, for example, to provide a statement about QoE with a particular communication service with known levels of QoS.

2.3 Classification of Methods for Quality Evaluations

Since it is of particular difficulty defining and assessing Quality of Experience, there are several methods to measure it. Choosing one rather than other is a matter of tradeoffs, depending also on several factors, like the experimental conditions. I here adopt the classification of U. Engelke and H.J. Zepernick that reported in their work [15] the classification of methodologies used to measure QoE, especially in the context of multimedia applications (e.g., voice- and video-streaming).

Subjective Methods

The evaluation of quality may be divided into two classes, subjective and objective methods. Intuitively, one can say that the best judge of quality is the human himself. That is why subjective methods are said to be the most precise measures of perceptual quality and to date subjective experiments are the only widely recognized method of judging perceived quality [16]. In these experiments humans are involved, having to vote for the quality of a medium in a controlled test environment. This can be done by simply providing a distorted medium of which the quality has to be evaluated by the subject. Another way is to additionally provide a reference medium, which the subject can use to determine the relative quality of the distorted medium. These different methods are specified for television sized pictures by ITU-R [17] and are, respectively, referred to as single stimulus continuous quality evaluation (SSCQE)[18] and double stimulus continuous quality-scale (DSCQS)[18]. Similar, for multimedia applications absolute category rating (ACR) and degradation category rating (DCR) are recommended by ITU-T in [19] and [20]. Common to all procedures is the pooling of the votes into a Mean Opinion Score (MOS) [20], which provides a measure of subjective quality on the media in the given test set.

Objective Methods: Psychophysical and Engineering Approach

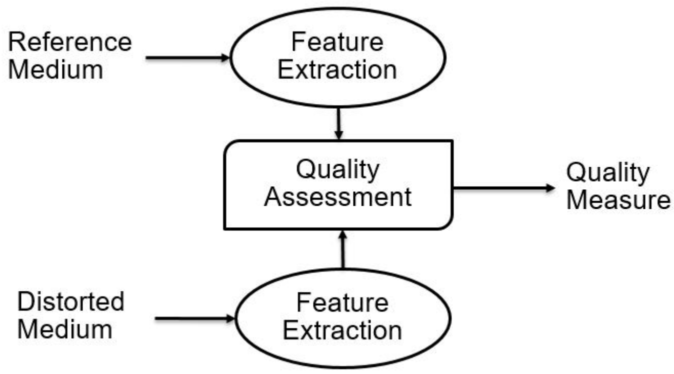
Two general approaches have been followed in designing objective quality metrics. They are often referred to as the psychophysical approach and the engineering approach. For the former, this can include modeling of contrast and orientation sensitivity, spatial and temporal masking effects, frequency selectivity and color perception. Due to the complexity of the human visual and perceptual system, these models, and their metrics, can become very complex and computationally expensive. On the other hand, they usually correlate very well with human perception and are usable in a wide range of applications. Methods following the engineering approach are primarily based on image/sound analysis and feature extraction, which also takes into account aspects of the human visual and perceptual system. The methods range from simple, numerical measures [21] to more complex extraction and analysis algorithms. The extracted features and artifacts can be of different kinds such as spatial and temporal information, codec parameters, or content classifiers. Simple methods are based on measuring single features whereas more complex algorithms combine various measures in a meaningful way. In any case, the metric outcomes can be connected to human visual and ear perception by relating them to MOS obtained in subjective experiments. Objective quality measurements can be further categorized into either "intrusive" or "non-intrusive" type, depending on whether they require some reference signal or not. Intrusive methods cannot be implemented in a real communications environment, because the reference signal is not available at the receiver side.

Full-Reference, Reduced-Reference and No-Reference Methods

Finally, one can classify quality metrics regarding their dependency on available reference information at the quality assessment equipment. The different methods that will be discussed are shown in Figure 2.1.

In general, it is no problem for the human visual and perceptual system to judge the quality of a distorted medium without having any reference available. However, it is a highly complex task for a machine. Metrics dealing with the approach of judging perceptual quality only based on the distorted medium are called no-reference or "blind" methods (Figure 2.1(a)). These approaches are easily applicable in a communication system as they reference quality prediction on the received data/media only. The task of no-reference quality assessment is very complex as no information about the original, undistorted medium is available. Therewith, a no-reference method is an absolute measure of features and properties in the distorted medium, which have to be related to perceived quality. The no-reference standard is the ITU E-Model, defined in the ITU-T Recommendation G.107 [22]. In order to quantify whether a change in quality between a reference and distorted medium has occurred, some degree of knowledge about the original medium would ease the related evaluation compared to using an no-reference method.

(a) No Reference method



(b) Reduced Reference method

(c) Full Reference method

Figure 2.1: Quality Assessment methods

This can be achieved by reduced-reference (RR) methods (Figure 2.1(b)). Here, only a set of features from the reference medium is needed at the quality evaluation equipment instead of the whole medium itself. This set of features can then be transmitted within the medium or over an additional channel. At the receiver, the features can then be extracted from the medium and used along with the reference features for quality prediction.

In all the cases where the reference is available at the evaluation equipment, one can use a full-reference (FR) method (Figure 2.1(c)). These methods use the reference to predict the quality degradation of the distorted medium. In general, this kind of approach could facilitate the process and it provides superior quality prediction performance. The existing metrics following the psychophysical approach are FR methods. PESQ [23] and POLQA [24] are the most used FR standard for speech quality assessment.

2.4 Elements of QoE in 5G Network-Design

As stated by C. Tselios and G. Tsolis in [25], the forthcoming advent of 5G networks will bring an ecosystem of applications that is expecting to improve most of the current characteristics of the legacy cellular systems, with increased levels of user experience. For this reason, elements of QoE have been taken under consideration by 3GPP since the early days of network design. The most prominent ones are summarized in this section.

2.4.1 Seamless Connectivity

5G UEs need to support a huge variety of both hardware and software technologies to provide consistent and uninterrupted service with a superior quality. Devices should be designed to face unpredictable issues related to channel conditions, high density of nodes per cell and the frequent needs of spectrum resources. From network operators side, the seamless connectivity could be achieved by a self-healing network architecture able to locate additional nodes, regardless of the failure reason, by tuning the operating channels over neighboring cells and by restoring the end users connectivity. The enhanced QoE-aware monitoring mechanism provided by NWDAF might prove to be useful on identifying compromised components, proactively redirect traffic and issue an alert towards the central network monitoring entity or simply save the particular report to the UDM, the integrated database function in 5GC.

2.4.2 Customized Service Distribution

In a user-centric communication network, NSSF is expected to correlate all the available services with a pre-existing or dynamically generated user profile. Especially when business models are involved, delivering personalized content through Network Slices, using different network resources per session, might be a necessity rather than a mere enhancement.

2.4.3 Neat Operability

5G network has been designed from its inception to deliver an improved level of services while preventing end-users being aware of functionality, error handling, resource allocation and traffic management. Interactivity between the user and the network should be limited to the absolute minimum. NWDAF, along with NSSF and PCF could provide quality feedback to the network itself, which will allow users to get benefits from a QoE-aware network. Obviously, data collection must be conducted in a fully automated manner, thus avoiding any level of annoyance to the end-user.

2.4.4 Energy Efficiency

Energy efficiency is amongst the areas that will undergo major redesign for meeting the high requirements of 5G. Several services will be moved into the network operators' domain, rather than maintaining them in the UEs - an approach called "edge offloading". With edge offloading, uplink and downlink decouple, whereas UE will be able of utilizing channels from different Mobile Broadband Systems. This will likely increase end-user satisfaction, through extended UE's battery life.

2.4.5 Virtualized Ecosystem

Recent advances in mobile cloud computing infrastructure allow scalable, on-demand access to a vast pool of configurable resources like processing speed, storage, networking and integrated applications over the Internet. This centralized operational model might improve end-users QoE as it reduces cost, increases availability, decouples services from the existing technology and offers flexibility in terms of provisioning.

2.5 Challenges for QoE Management in 5G

In this section I am going to summarize the salient aspects that will bring new challenges to QoE management. These include explosive growth of data traffic, massive increase in the number of interconnected devices, rising of new services and application scenarios [26].

2.5.1 Heterogeneous Radio Coverage Scenarios

In 5G, typical scenarios include communications when people are at work, when people are at home or entertainment events, and when people are on the move [27]. A seamless QoE is expected in all types of radio coverage scenarios, including in buildings, dense metropolitan areas, rural areas, stadiums, subways and highways. These scenarios, which are characterized by ultra-high traffic volume density, ultra-high mobility, are challenging for QoE management in 5G Networks. For instance, in a high-speed scenario, maintaining a satisfactory level of service continuity is of great importance. Similarly, designing an appropriate mobility management strategy in the ultra-dense small cells scenario is fundamental to reduce the number of handovers, which is a key issue to improve QoE. Therefore, maintaining a consistent and acceptable user experience in such heterogeneous network settings requires a tight coordination of the various QoE control strategies.

2.5.2 Emerging Applications

With the rapid development of computer hardware and software as well as the standardization of Internet technology, intelligent mobile terminals and smart wearable terminals have various forms (i.e., watches, glasses and sports wristbands, just to name a few). At the same time, mobile services, which are based on smart terminals, are no longer limited to cell phones, text messaging, and video services; rather, they play a key role in medical monitoring, interactive games, information exchange, and other emerging verticals. Since most existing QoE models focus on VoIP, video-streaming and HTTP services, the new service characteristics and user demands should be expanded to setup a proper QoE model for the above-mentioned and recently developed applications.

2.5.3 Energy Demands

As stated before, energy efficiency is another area that will experience major redesign for meeting the high requirements of 5G Networks, at it could have a great impact on the overall QoE. With new attractive applications, users will use their equipment and mobile phone, and more frequently. Consequently, the terminal should have a long enough battery life to ensure reliable service even with more extensive usage, otherwise QoE may drop significantly. As noticed by Andrews et al. in [29], continuous increase of power consumption is not viable from logistical, cost and battery-technology perspective. Rapid increase of network density is directly linked to elevated energy demands on the radio side, while the necessity of UE for seamless connectivity and expected support of a broad spectrum of new applications, software and utilities increases the overall computational cost and diminishes the average battery duration per charge. M2M communications (under the Internet of Things umbrella) are another areas that can be considered energy-sensitive, since wireless sensors often have limited operational capabilities due to inefficient batteries. Furthermore, the introduction of NWDAF, a QoE-aware mechanism able to obtain network datasets in a real-time way, as well as to utilize them in proactive network management, will add significant end-to-end complexity. For this reason, all the network functions should extend energy provisioning for the expected control signaling overhead, while, at the same time, trying to retain the standards of service to the highest possible level.

2.5.4 Handling Massive and Heterogeneous Amount of Data

In the 5G era, with the exponential growth in network data traffic and the rising of heterogeneous applications, it can be easily predicted that huge amounts of data will be generated. This trend will likely rise several challenges in the fields of QoE management. Firstly, the challenge related to capture the subjective factors affecting QoE, for instance understanding the users specific preferences in the massive and diversified end-users expectations scenarios. As a matter of fact, there are several subjective elements that are needed to be taken into account, such as users mood, attention, expectation [28]. In this regard, one should be able to answer to a series of problems involved with these factors, like how to quantify and subsequently normalize them, or how to capture them in real-time applications. Furthermore, assessing QoE in the era of security and privacy related to intelligent terminals is another challenging matter. This comes along with the trend that smart devices play a non-negligible role in 5G: these devices could collect users personal information, contacts, download history, application usage records, and system logs, thus inferring a users own personality traits and preferences, consumption habits, and even values, in order to provide personalized QoE. Since personalized service involves the collection and analysis of a users personal information, the greatest challenge comes from balancing the protection of a users privacy and, at the same time, ensuring personalized QoE management.

2.5.5 Over-The-Top (OTT) Operators and Encryption

Currently more than 60% of mobile traffic is encrypted, a trend that is rapidly rising [29]. Mobile video will increase 11-fold by 2020, accounting for 75% percent of total mobile data traffic [30]. Such rapid growth impacts on network operators who have to thoroughly change and optimize their network. In particular, control and maintaining satisfactory QoE for multimedia streaming services is becoming a greater challenge for network operators than ever before [31]. As early mentioned, downloading and watching video content on mobile devices is currently a growing trend among users. To perform capacity planning, network operators have to deeply understand and track the offered

QoE on multimedia content delivery. Techniques such as caching, transcoding, compression and radio resource allocation across users have been introduced to facilitate the delivery of media-rich content. However, at the same time, a significant number of major Internet services have begun to protect and encrypt their traffic. Popular OTT video providers such as YouTube, Netflix and Hulu now encrypt a large part of their video content. This trend suggests that most of multimedia traffic will be encrypted in the next years [32]. Network operators are blind towards services delivered by OTT, thus compromising their ability to monitor and keep track of the overall experience they offer to end-users [33].

Chapter 3

VoLTE: A Promising 5G Application

LTE and LTE-A are the latest fully standardized technologies for cellular connectivity and their evolution poses new basis to the commercial advent of 5G. As previously remarked, 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:

1. Circuit Switched Fall Back (CSFB) [34], that relies on the preexistent GSM/UMTS networks;

2. VoLTE via IP Multimedia Subsystem (IMS), defined by GSMA in 2014 [35], 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, 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 3GPP and 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 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, I leveraged 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.

The remainder of this chapter is organized as follows. Section 3.1 gives an overview of the existing contributions. Section 3.2 depicts the background for this work, touching upon VoLTE architecture, currently adopted voice codecs and objective voice quality assessment. Section 3.3 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.

3.1 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 types of service, such as real-time communications and streaming applications [36]-[37]. Within this framework, in [36] 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 [38] 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 [39]; in detail, a comparison was set up, to confront VoLTE against circuit-switched and Skype/Google Hangouts voice calls. In [40], 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 [37], the dependency of the average VoLTE call duration on call quality was investigated. Finally, in [41] 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 my research 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.

3.2 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.

3.2.1 VoLTE Architecture

Fig 3.1 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), are 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 (QCI) [42], 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.

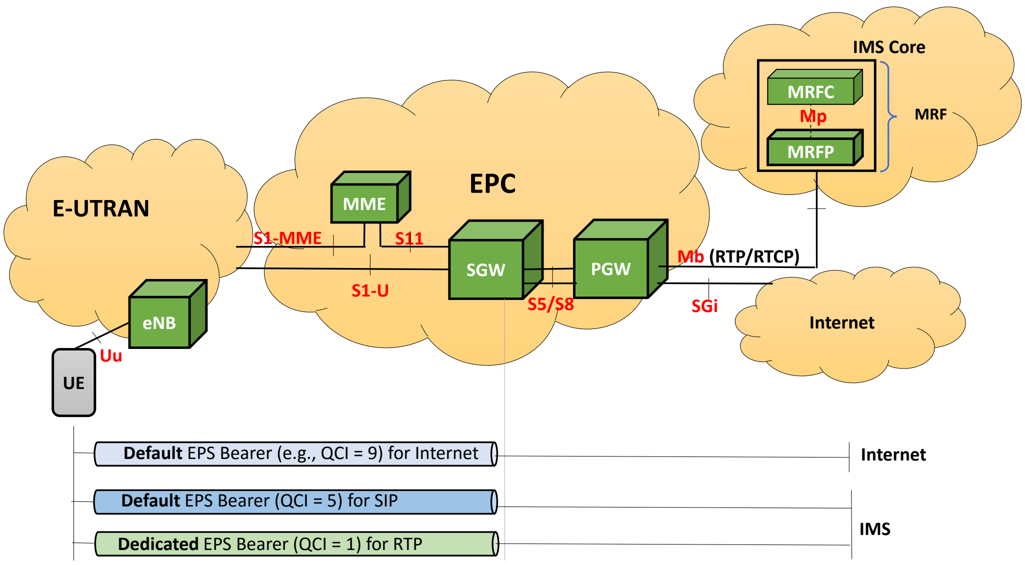


Figure 3.1: Simplified LTE and IMS network architecture

3.2.2 Codec Overview

When it comes to the voice codecs most widely adopted in current days, it can be observed that the Adaptive Multi-Rate (AMR) [43] codec is a narrow-band codec largely popular in GSM and UMTS. It was originally developed by 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 [44]. The AMR encoder is able to dynamically adapt its output rate to the current radio channel conditions, featuring 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 Wide-Band codec (AMR-WB) [45] 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 [46] and [45] 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"*.

3.2.3 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. As already outlined in chapter 2, 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 for 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® [47]-[48] is the objective, non-intrusive tool employed in this study; it is an extension of the the E-Model [22], 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.

Table 3.1: Classes Definition of Speech Transmission quality

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

The R-factor can be suitably mapped to the Mean Opinion Score (MOS)[20] on the well-known $1-5$ scale. Such correspondence has been recently updated for the Wide-band version of the E-Model [49], where the R-factor can reach values up to 129, as it might happen when the AMR-WB codec is used. Table 3.1 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 affects 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.

3.3 Settings and Results

3.3.1 Data Collection

I 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. This probe, purchased by Empirix, has an Intel® Xeon® Processor E5-2680 v2 with $192GB$ RAM and $4TB$ x 12 as HDD.

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, I 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 my 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, I was able to estimate the packet loss rate, evaluating the ratio between the number of lost or 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 [50] or alternative, less popular speech audio codecs, I was 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 I based my analysis, first focusing on LTE network conditions, as discussed next.

3.3.2 Network Conditions

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

Fig. 3.2 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 0.01 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. 3.3 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. 3.4(a) and (b) provide a unified view of the examined LTE network: Fig. 3.4(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

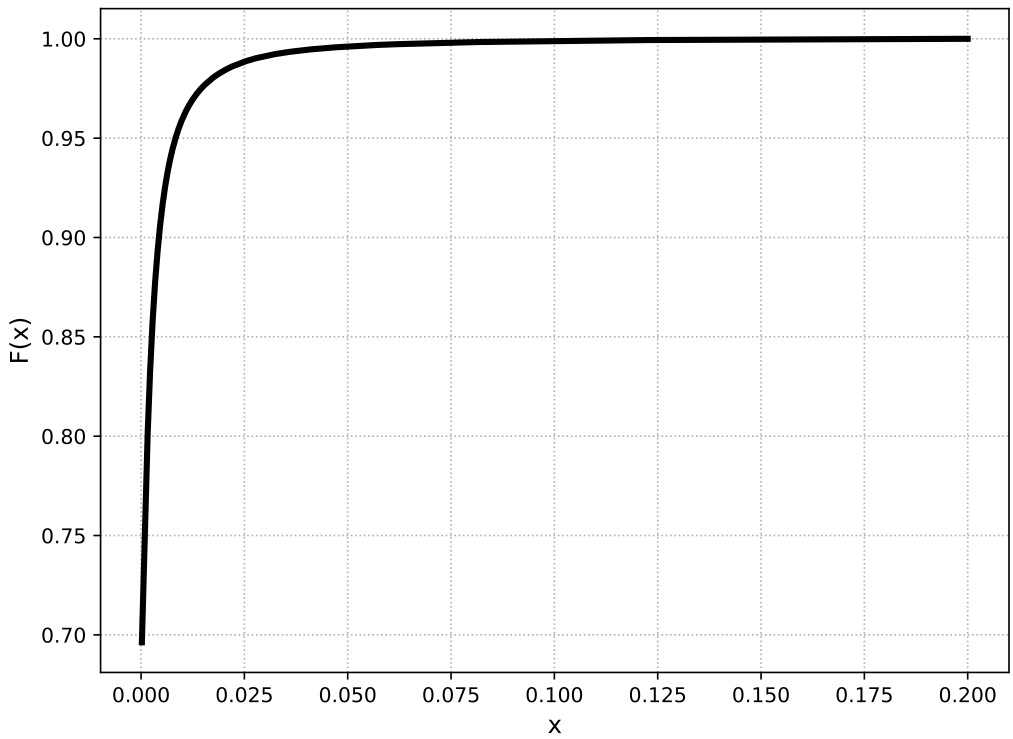


Figure 3.2: Packet loss rate CDF

based voice flows undergo, whereas Fig. 3.4(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 experience in LTE, when clients are equipped with a playout buffer.

3.3.3 R-factor Results

Next goal was to investigate the QoE perceived by VoLTE calls. To this regard, Figs.3.5(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.3.5(b) is exclusively due to the lack of points in the region of high packet loss rates and high values of maximum

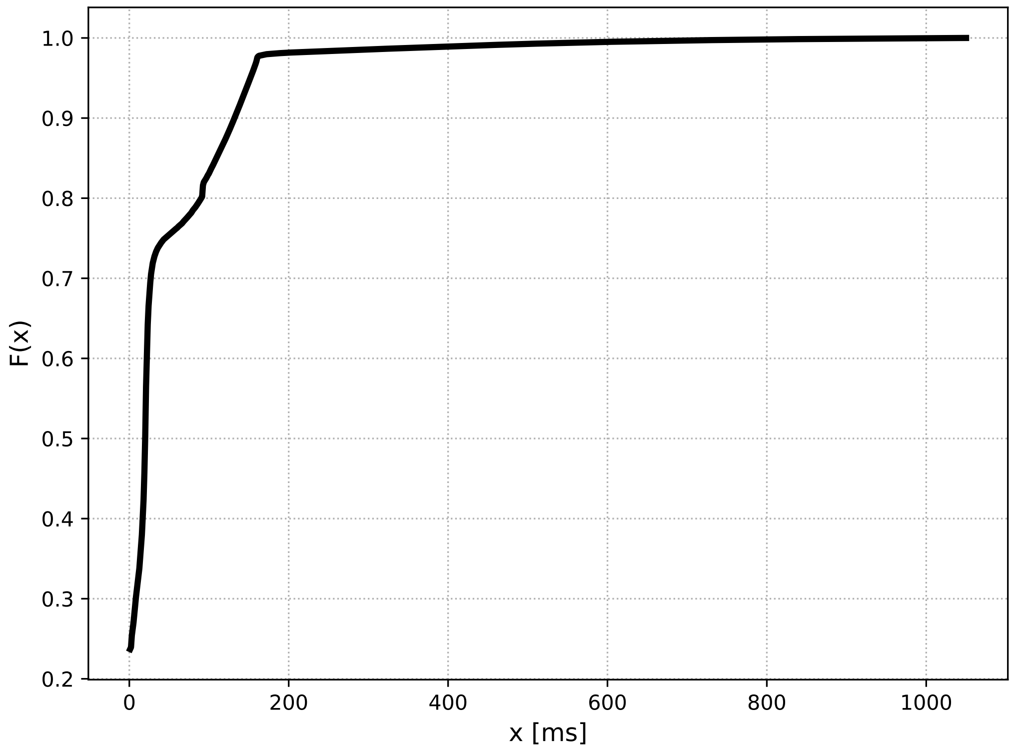


Figure 3.3: CDF of the maximum jitter

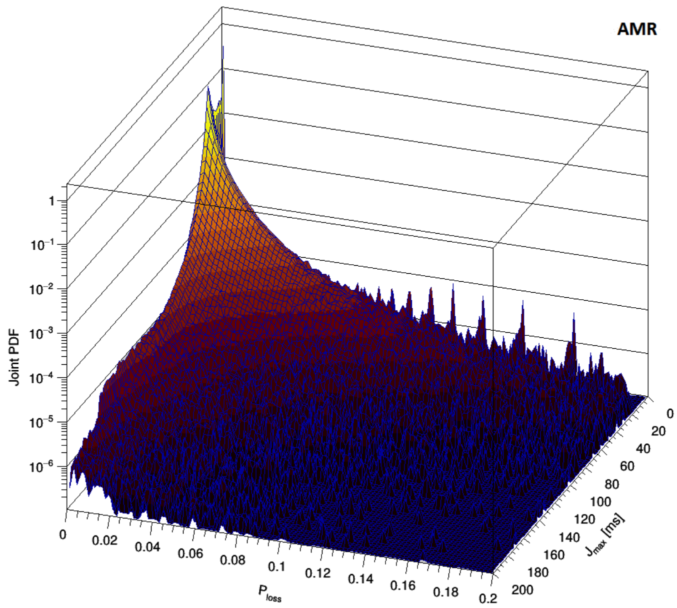
jitter, that the former Figs.3.4 (a) and (b) already evidenced.

In order to better investigate the R-factor dependency on P_{loss}, Fig. 3.6 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, 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, $i = 1, 2, \dots, 10$, 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, I conclude that the fitting is truly satisfying. To confirm the goodness of the exponential choice, I 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, I 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.

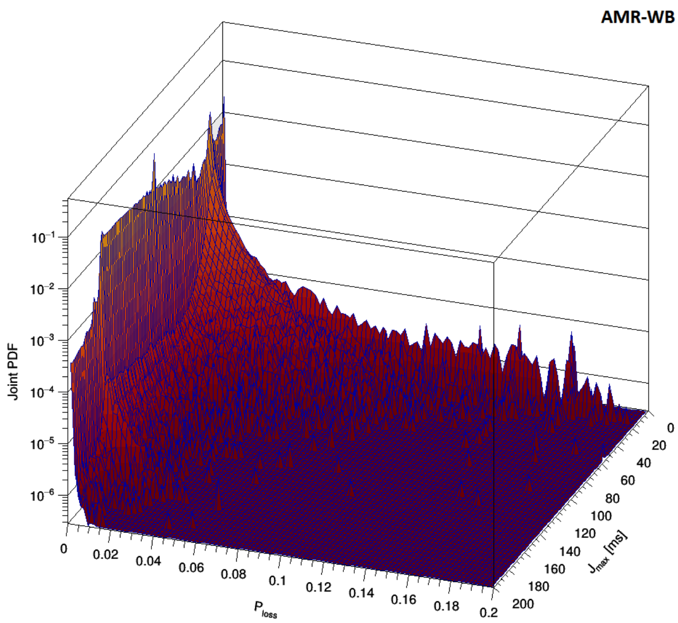
For the AMR-WB case, Fig.3.7 reports the original (P_{loss}, R) points, together with the comparison between the exponential and the linear fitting. For the former, $y_0 = -31.72$, $A = 135.07$ and $B = 0.28$, whereas for the linear regression $A' = 99.01$ and $B' = -340.7$. Interestingly, in this case I 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.

Although not reported in this chapter, I 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 [36].

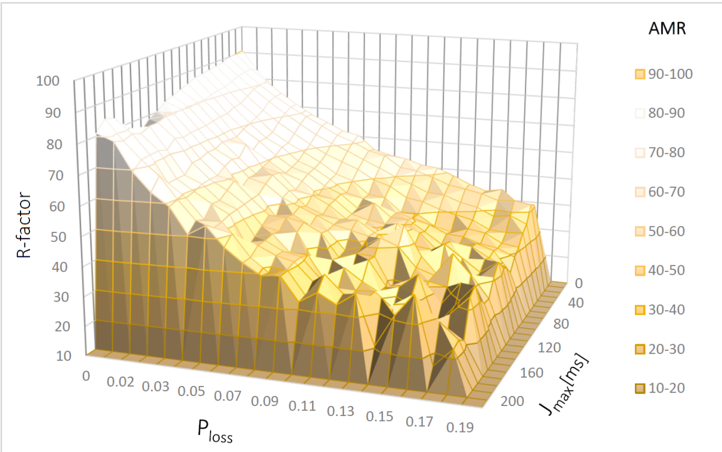


(a) AMR based flows

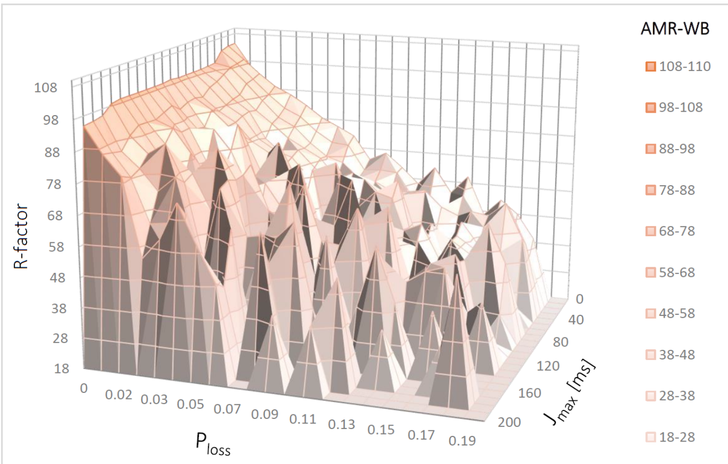


(b) AMR-WB based flows

Figure 3.4: 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

Figure 3.5: R-factor as a function of the packet loss rate P_{loss} and of the maximum jitter J_{max}

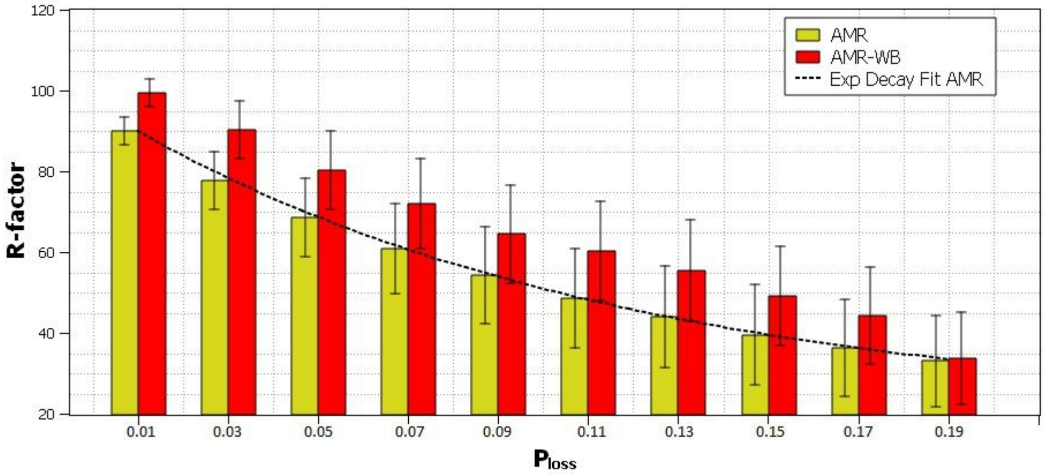


Figure 3.6: R-factor as a function of P_{loss}

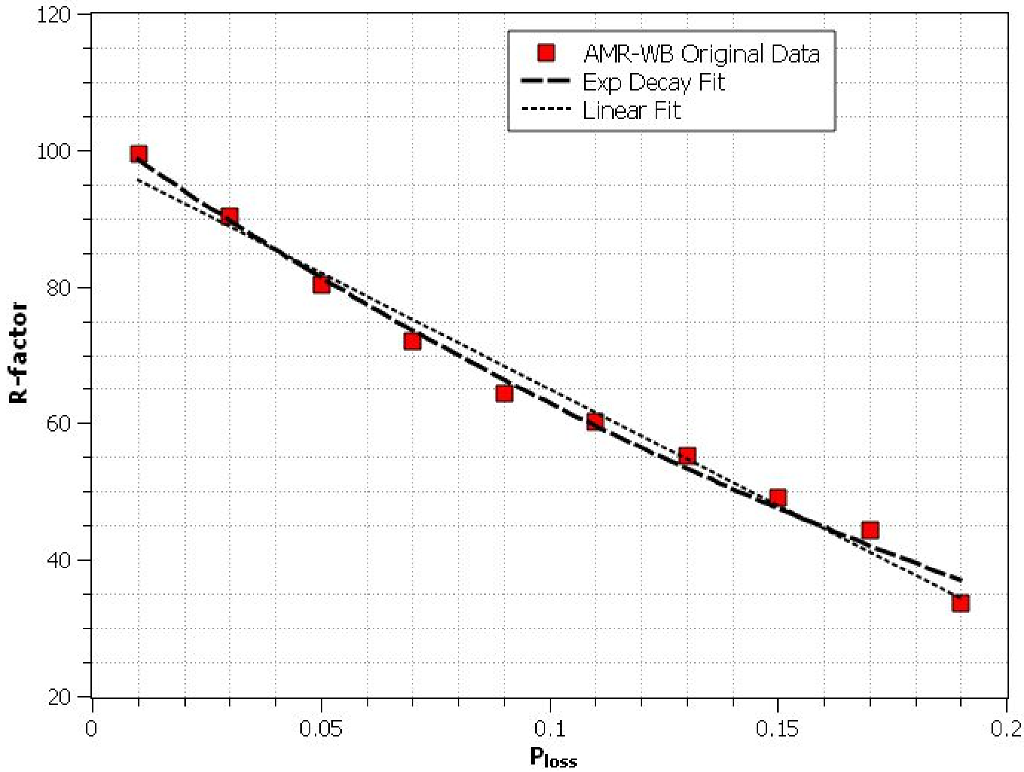


Figure 3.7: Fitting comparison for the AMR-WB case

Chapter 4

A Novel Approach for Speech Quality Assessment

4.1 Introduction

Given a certain applications, one of the main goal of a Service Assurance company, like Empirix, is to get a QoE estimation out of a set of measurable input parameters. In a more formal view, this can be seen as the general problem of modeling a mapping function f to assign QoE values from a set of measurable parameters $P_1, P_2, ..., P_n$ as represented in Figure 4.1.

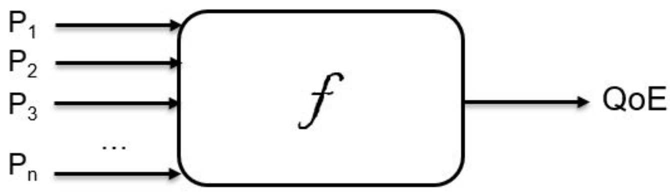


Figure 4.1: Map of QoS/QoE general problem

$P_1, P_2, ..., P_n$ could be either QoS Network Features (e.g., Average Jitter, Packet Loss Rate, ...) and/or other factors that influence QoE (e.g., environment, mood, ...). Finding the function f that relates the input parameters with QoE is one of the most challenging problems the scientific literature has been investigated since years. In this view, Machine Learning (ML) can be a powerful tool to discover the hidden correlation (i.e., the function f) between network metrics and the user experience, directly from the data.

A research field at the intersection of statistics, artificial intelligence, and computer science, ML, also known as “predictive analytics” or “statistical learning”, deals with extracting knowledge from data. Even though ML turned out to be a buzz-word, it is important to recognize that its methods of application have in recent years become ubiquitous in everyday life. From automatic recommendations of which movies to watch, to what food to order or which products to buy, to personalized online podcasts and recognizing persons in photos, many modern tools and devices have machine learning algorithms at their core. Thus, given the huge amount of data traffic that 5G networks will generate, and the emerging customer-centric technology, ML is expected to exhibit the right requirements to approach the problems related to QoE. Once a ML-based model has been properly trained, it could give any enterprise the opportunity to identify problems and their causes, saving operational costs.

To this end, Empirix and I have leveraged ML techniques to explore novel ways to assess QoE in the promising context of voice delivery applications ensured by 5G. As a matter of fact, QoE of VoIP calls is a relevant topic in the realm of contemporary networks, given VoIP widespread adoption in wired scenarios, but even more in cellular networks, where its counterpart, VoIP over LTE (VoLTE), combined with super-wideband codecs, plays the leading role in ensuring high quality levels to voice calls in a totally IP-based scenario. In a previous study [51] - [52], discussed in chapter 3, I assessed the end-to-end transmission quality of several millions of VoLTE calls employing VQmon®[48], an objective, non-intrusive tool, that enhances the standardized E-Model[22]. Tools like VQmon® are quite popular on the service assurance rim, as they can be easily integrated in proprietary software. Yet, they are quickly becoming obsolete, given the complexity and heterogeneity of modern communication systems [53]. In particular, VQmon® puts the emphasis on packet loss rate, and my focus is also to explore the effects of jitter and delay on QoE.

Taking these remarks as its starting point, the aim of this chapter and of my most recent research work [54] is two-fold:

1. first, to quantify VQmon® limits in QoE assessment of VoIP calls that employ a wideband voice codec;

2. then, to overcome such limits proposing the adoption of a supervised ML approach.

With reference to the latter point, the current study demonstrates to what extent Ordinal Logistic Regression (OLR) performs better than other popular state-of-the art ML solutions. It therefore proves that the OLR algorithm is well suited to model the human level of preferences expressed on an ordinal rating scale.

In order to achieve the goals stated above, a subjective listening campaign has been led in a controlled environment; the transmission of wideband, high quality VoIP calls, has been repeatedly mimicked, collecting network metrics and several categorical features of the volunteers participating to the quality assessment test. Participants have been asked to rate the listening quality of test calls and the test outcomes have first of all disclosed VQmon® flaws. Most importantly, they have allowed to highlight the benefits of the proposed ML approach, which is fast like non-intrusive methods, as it automates speech quality prediction, and trustable, being built on a subjective basis that can be retrained several times upon customer availability and network adjustments. My proposal therefore embraces the cause of mobile operators and network monitoring companies, that not only mandate for effective monitoring tools, but also for easiness of deployment on millions of VoIP calls.

The study further highlights that the conventional five score scale for call quality classification is often perceived as excessive by test participants. In the limit case where ratings are collapsed on a coarse binary scale, OLR and alternative ML models are verified to guarantee a very high and comparable accuracy level.

The remainder of this chapter is organized as follows. Section 4.2 summarizes the existing contributions. Section 4.3 gives an overview of the employed ML models and of the experimental environment. Section 4.4 illustrates the data collection process and then discusses the results obtained in terms of performance prediction.

4.2 Related Work

In the past, a few solutions based on advanced statistics and ML models such as Bayesian Classifier [55], Artificial Neural Networks [56] and Random Neural Networks [57] have been proposed to predict VoIP speech quality. As a recent example belonging to this category, the study in [58] compares the performance of different ML classifiers, considering packet loss, narrow-band codec type, language and gender as features. All the previously cited works assume as learning basis (equivalently termed ground-truth) the quality ratings that the Perceptual Evaluation of Speech Quality (PESQ) technique [23] provides. PESQ is an algorithm for narrow-band voice evaluation; it is objective, i.e., it automatically evaluates speech quality with no involvement of human subjects, and it is double-reference, as it compares the received voice signal against the clean, original signal. However, one relevant drawback inherent to the choice of employing PESQ outcomes as ground truth is that the estimate error affecting the reference technique propagates to the learning algorithm. Alternative studies, like [59], considered as ground-truth the subjective Mean Opinion Score (MOS), that ITU defines as the arithmetic mean of a collection of single user opinion scores [20]. Yet, the arithmetic mean might represent a rough approximation when judging the quality of VoIP calls: it inevitably smooths out the quality score that a specific user assigns the call under certain network conditions. Lastly, P. Charonyktakis et al. [60] designed a modular algorithm that uses multiple ML models, including Decision Trees and Support Vector Regression, and relies on an optimized technique, termed nested cross validation, to select the best classifier. This study adopts both subjective tests and PESQ to rate the actual QoE of narrow-band VoIP calls.

Partly in analogy to the contribution in [60], this study concentrates on the subjective experience of single users as ground truth. Differently from [60] and previous works, our study proposes to handle the rating of the call quality experienced by the single user as an intermediate problem between regression and classification. It therefore suggests to exploit a specific algorithm, the so-called Ordinal Logistic Regression (OLR), and it benchmarks its performance against some of the most popular ML methods already utilized in the works cited above, highlighting its better accuracy. Further, my contribution concentrates on wide-band, high definition voice, which is of paramount importance in VoLTE, as well as in 5G networks. To the best of my knowledge, all the investigations on VoIP QoE presented so far in literature are centered on the adoption of narrow-band codecs, that work on audio frequencies in the 300-3400 Hz range. However, all modern applications relying on telephony audio employ wideband and super-wideband codecs, which extend the maximum operating frequency to 7 and 22 KHz, respectively. This is the case I therefore choose to concentrate on.

4.3 Background and Setting

This section is intended to provide a brief overview of the employed ML models and of the experiment setting and design. An extensive explanation of the selected algorithms and of their implementation details can be found in [61] and [62].

4.3.1 Prediction Models

The distinctive feature of supervised learning is that the target variable to predict is known (e.g., in this work I know the QoE labels), and this information is explicitly used in the learning process. Moreover, supervised learning approaches can be distinguished in classification and regression solutions. I refer to classification when the target variable is a class, as in the examined problem. Decision Tree classifier (DT) is a ML classification algorithm that produces interpretable models and it is widely employed for this distinctive feature: its goal is to create a model that learns from simple if/else rules inferred from data. To build a tree, the algorithm searches over all possible paths and finds the one that is the most informative about the target variable. An enhancement to DT is Random Forest (RF), fitting a number of DT classifiers on various subsets of the dataset. RF relies upon an ensemble of trees to improve predictive accuracy. Trees can be easily visualized and interpreted, but their main drawback is that they neglect any ordered trait of the target feature. Differently from classification, regression predicts a continuous outcome. The reference model is Linear Regression (LinReg), utilized to find the relation between two or more continuous variables. Logistic Regression (LogReg) replaces LinReg when the target is no longer continuous and is expressed as a dichotomous variable. Its generalization to more than two classes is Multinomial Logistic Regression (MLR). Lastly, OLR represents an intermediate approach between classification and regression, and it is my belief that it can successfully fit the present problem of predicting *ordered* classes of QoE. As a matter of fact, OLR handles labels that are both discrete as in classification, and ordered as in linear regression. Its complex mathematical formulation is based on the generalized linear model, well-detailed in [63] and [64].

4.3.2 Experiment Setting and Design

Fig.4.2 portrays the end-to-end setting of the experiment. Calls were generated by Hammer®, a proprietary platform by Empirix [65] that emulates software agents initiating and accepting VoIP calls and establishing an SIP/RTP session for every call. One Hammer was installed on the Virtual Machine (VM) of a Windows PC, acting as the caller (Hammer A), a second Hammer was installed on the VM of a second PC, representing the callee; a Linux-based, Ubuntu VM on a third PC routed packets from the caller to the callee, and also acted as a source of impairments through Netem[66], a network simulator available in Linux kernels. All PCs belonged to the same Gigabit Ethernet LAN.

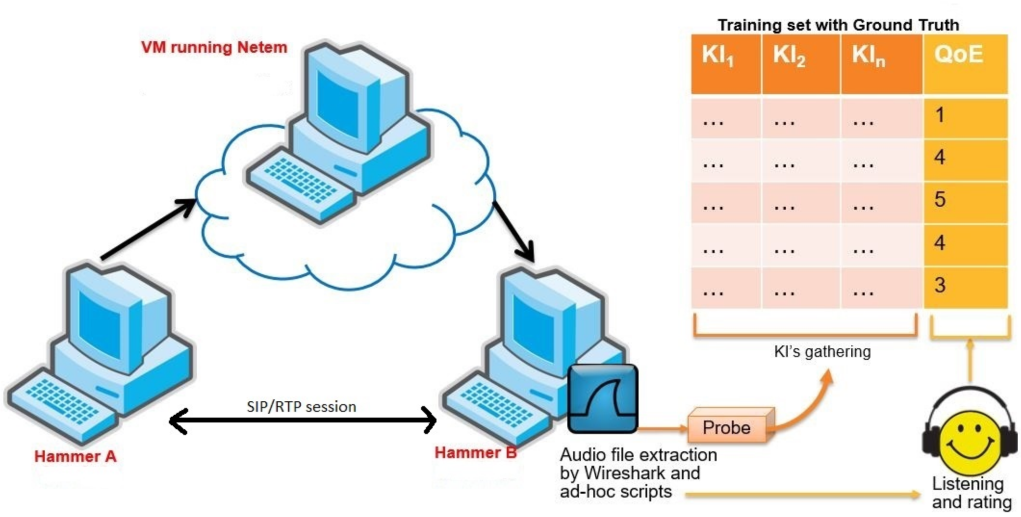


Figure 4.2: End-to-end experiment workflow

I chose to deliver the short audio stream "You will have to be very quiet", encoded through Adaptive Multi Rate WideBand (AMR-WB) [45] (mode 25.85 kb/s) and fully compliant with ITU-T guidelines about subjective listening tests [20]. Each call featured the same audio stream. Through Netem[66], I intervened on the one-way delay and packet loss to simulate the typical impairments of real networks. In detail, given the ITU-T G.1010 document [67], that suggests the tolerated values of one-way delay (lower than 400 ms) and packet loss rate (lower than 1%) for conversational audio, I combined four profiles of packet loss (random, uniformly distributed losses with rates 0%, 0.5%, 1% and 2%) with three profiles of one-way delay (Gaussian distributed with mean and standard deviation equal to (0 ± 0)ms, (150 ± 25)ms and (400 ± 25)ms), thus obtaining twelve scenarios. At callee side, the jitter buffer was instantiated to receive packets with a fixed inter-packet delay. The received files were collected in a Wireshark [68] compatible format, and sent to a proprietary probe, where they were processed and then exported. Since I operated in a virtual environment, I made use of *ad-hoc* scripts[1] to extract the audio trace in a listenable format.

I next conducted a subjective listening campaign, and designed the listening experiments in accordance to ITU-T guidelines [20]. Among the available quality assessment methods, I adopted the popular Absolute Category Rating (ACR) test, because of its reliability and fast implementation. In ACR subjective tests, users are asked to evaluate calls, presented only once, and have to rate the listening quality, i.e., their QoE, on an ordered scale featur-

[1] `https://github.com/Spinlogic/AMR-WB_extractor`

ing five score values: 1 (bad), 2 (poor), 3 (fair), 4 (good) and 5 (excellent). For the experiment, a pool of 56 participants was recruited on a voluntary basis in the first half of 2019. Every listener was asked to evaluate the quality of 12 calls, corresponding to the received audio streams in the 12 scenarios described above. Before starting the survey, I additionally asked volunteers to answer a few questions, namely, to indicate their gender, age and the type of headset employed during the test. These features uniquely characterized each participant, along with the rating she/he attributed to the quality of the calls. In addition, I encouraged users to share their feedback. At the end of the experiment, I collected a total of 672 evaluations. Because of the arbitrary property of subjective tests, it is known that some ratings might have been assigned in an inappropriate manner. Thus, I grouped call scores by call identifier and applied the popular DBscan algorithm [69] to detect outliers among the evaluations collected for each call. DBscan found a total of 55 outliers, that were removed from the initial set.

4.4 Experimental Results

4.4.1 Data Set Preprocessing

The dataset available after the listening campaign included the actual network metrics characterizing each evaluated call, that is, the following numerical features: average and maximum jitter, number of received packets, packet loss rate, out-of-sequence packets and duplicated packets, as well as the set of categorical features directly collected from the participants, that is, their age, gender and type of headset. Most importantly, I collected the rating each participant attributed to the quality of the calls, i.e., their QoE scores.

To minimize the risk of injecting noise in the model, I firstly determined the most informative features with respect to the target label, i.e., the QoE score. According to Pearson correlation test [70], those numerical features exhibiting a p value greater than 0.01 were considered insignificant. I therefore neglected the number of received packets and the number of duplicated packets. Given the relatively modest number of examined settings for the test, I additionally "flagged" the packet loss rate as a binary variable: I stated that it was *present* in any scenario where it took on values greater than 10^{-2}, otherwise it was interpreted as *absent* (for the examined scenarios, this corresponds to values lower than 10^{-3}). Lastly, as the examined numerical features span on different scales, I rescaled them, in order not to privilege one over others (e.g., maximum jitter over out-of-sequence packets).

Given the relatively low number of research participants, I decided to include all the categorical features in the present study, as it is not possible to firmly state that QoE is independent of them.

4.4.2 Exploratory Investigation and Performance Assessment

Preliminarily, I investigated the reliability of VQmon® when assessing the quality of VoIP calls; for doing so, I compared the MOS values that VQmon® provides against those determined from the actual subjective ratings; adhering to MOS definition, I computed the latter value as the average of the individual ratings that different users assigned to the same call. Fig. 4.3

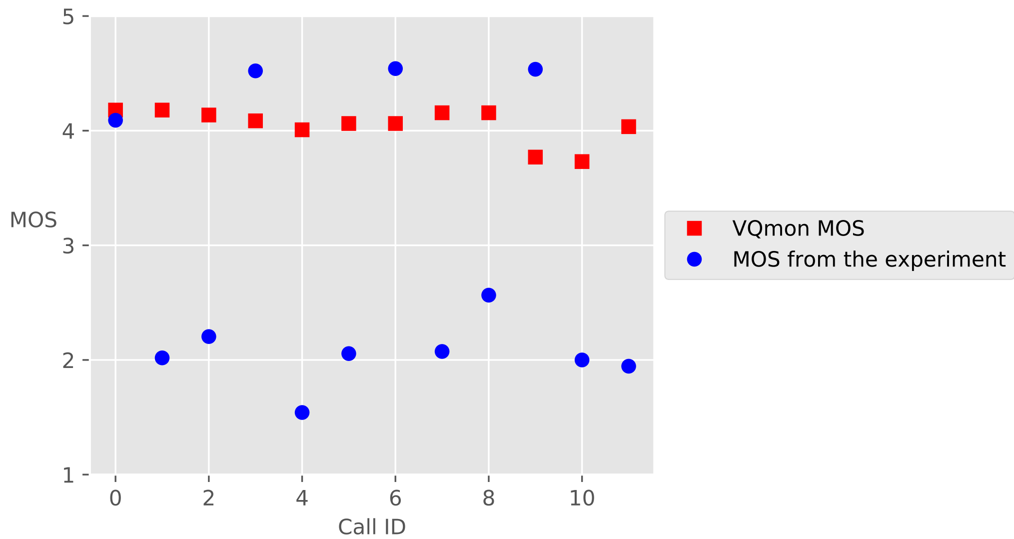


Figure 4.3: VQmon® MOS and Users MOS per Call ID

shows how far VQmon® MOS values (red squared markers in the figure) are from their experimental counterparts (blue circles) for the 12 synthetic calls evaluated by the users. The results reported in this figure clearly demonstrate that VQmon® cannot predict the actual call quality, and further motivates us to explore the effectiveness of a user-driven methodology that leverages ML tools.

Given the presence of ordered classes, i.e., the five possible QoE scores, I deliberately focused on OLR as a promising candidate among the alternative ML algorithms. To validate the goodness of such a choice, I considered a random split of the QoE scores, employing 80% of them as the training set and the remaining 20% as the test set and first benchmarked OLR classification accuracy against that of the Random Classifier (RC), DT, RF and MLR. Given the relatively few training data available, I decided to exclude neural networks from my investigation [71]. Moreover, I did not consider Support Vector Machines (SVM) models either, because they do not perform well with unbalanced classes [72], as it is the case here. Recalling that accuracy is defined as the percentage of correct predictions to the total number of test samples, I found that OLR outperforms all other algorithms. As a matter of fact, its accuracy is 61%, almost four times the RC accuracy, which amounts to 16%, and higher by ten percentage points than both DT (51%) and MLR (52%), where I emphasize that the latter two algorithms do not take into account class ordering.

Table 4.1: OLR and DT Confusion Matrix

		Predicted by OLR				
		1	**2**	**3**	**4**	**5**
Observed	**1**	2	15	0	0	0
	2	2	39	3	0	0
	3	0	13	6	2	3
	4	0	0	0	12	6
	5	0	0	0	3	17

		Predicted by DT				
		1	**2**	**3**	**4**	**5**
Observed	**1**	0	17	0	0	0
	2	0	44	0	0	0
	3	0	19	0	5	0
	4	0	0	0	4	14
	5	0	0	0	5	15

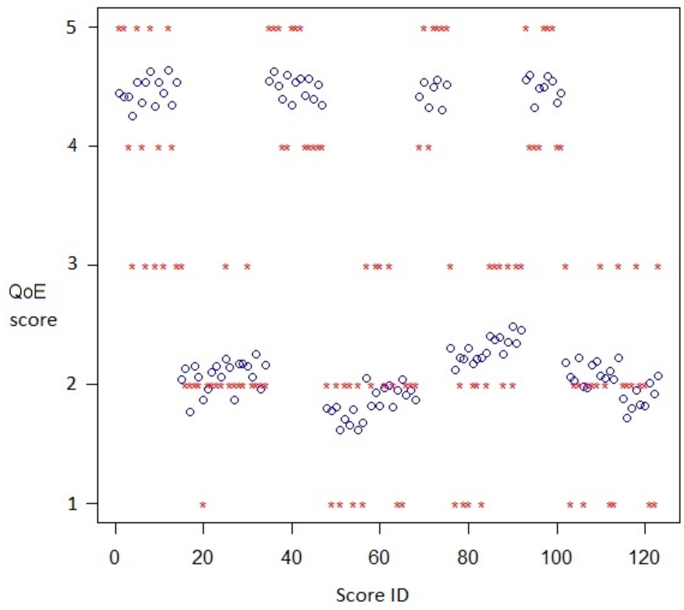
Confusion matrix (CM) generalizes the concept of accuracy: every CM row represents the instances in an actual class and every column represents the instances in a predicted class, so that the ideal CM has zero elements everywhere except for the main diagonal, meaning that all the predicted instances coincide with the actual observations. Table 4.1 compares OLR and DT confusion matrices, revealing that OLR better captures intermediate opinions (3 and 4 QoE values), that are more likely related to each test participant and her/his set of unmeasurable characteristics (e.g., mood, tolerance level), whereas DT limits its prediction to three out of five QoE classes (2, 4 and 5). To exclude that this study had to be approached as a linear problem, I further considered LinReg as an alternative baseline. I therefore extended the domain of the target label QoE from integer to real, thus removing the concept of classes. The subjective QoE scores (red crosses) and the predicted values (blue circles) are reported in Fig. 4.4(a) for LinReg and in Fig.4.4(b) for OLR. They allow to compare the performance of LinReg and OLR, revealing that LinReg is unable to predict intermediate results, whereas OLR can.

Lastly, it is interesting to outline that out of 56 research participants, almost half of them pointed out that five classes were too many to evaluate the QoE of the test calls, which might more simply be rated as poor or good. Adhering to this rationale, I remapped the five original classes into two, class 0 collecting the previous 1 and 2 classes, and class 1, merging classes 3, 4 and 5, so as to reduce the problem to binary classification. As such, the concept of ordering no longer holds, and the binary counterpart of OLR is LogReg. When taking this approach, both accuracy and confusion matrix remarkably improve, and as expected, LogReg and DT exhibit similar performance. In detail, LogReg accuracy stands at 83% and by inspecting LogReg and DT CMs reported in Table 4.2, I observe the prevalence of correctly predicted instances.

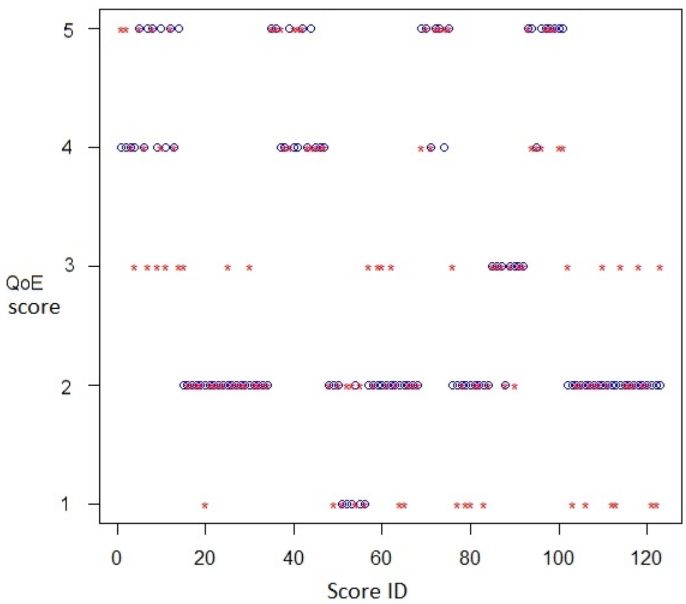
Table 4.2: DT and LogReg Confusion Matrix (Binary classification)

		Predicted (DT)	
		0	1
Observed	0	62	0
	1	21	41

		Predicted(LogReg)	
		0	1
Observed	0	60	2
	1	18	44



(a) LinReg



(b) OLR

Figure 4.4: LinReg and OLR performance

Chapter 5

Conclusion

In this thesis I focused on the increasing interest towards Quality of Experience research in the context of LTE and 5G networks, proposing a novel approach to assess the quality that end-users will undergo.

At first I have investigated the key aspects of 5G networks following the standardization process promoted by 3GPP and ITU. I then narrowed the focus of my research towards the new architectural concepts addressed to enhance the overall Quality of Experience, a strategical indicator for Empirix, given it has been identified as one of the salient differentiating elements with respect to the previous generations of cellular networks.

Next, I performed a comparative analysis of the end-to-end quality guaranteed by VoLTE, focusing on calls that employ two popular speech audio codecs, namely, Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wide Band (AMR-WB). The corresponding 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, I have 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.

At the end of this study, I have conducted a subjective campaign of quality assessment on artificially generated VoIP calls, collecting the values of network metrics associated to each test call, some categorical features of the participants and their QoE scores. This work has first demonstrated the shortcomings of a conventional objective, no-reference model when assessing speech quality of VoIP wide-band calls. Next, I have proposed to adopt a customer-driven, Machine Learning approach to correlate network-oriented features and human-related aspects to the levels of QoE that listeners perceive. Ordinal Logistic Regression (OLR) has been proved to be the best algorithm to model the examined problem, as it better approximates the ordinal behavior of subjective experience. The study has additionally provided an insight into the difficulties of utilizing a five level scale to evaluate VoIP QoE, often perceived by test participants as poor or good. When handling the quality assessment problem as binary instead of ordinal, I have shown that both Logistic Regression, the binary counterpart of OLR, as well as alternative ML algorithms (such as Decision Trees and Random Forest), guarantee reliable and similar predictions. From network operators side, the output of this research activity could help customer experience managers and service quality managers to identify in a fast way potential issues in their core network depending on predicted values of the QoE.

Future research work should aim at collecting additional measurements from access network. This would allow to improve the generalization capabilities of the OLR algorithm and to evaluate its performance under different operating conditions.

Bibliography

[1] ITU-R. IMT Vision - Framework and overall objectives for the future development of IMT for 2020 and beyond. *ITU-R M.2083-0*, 2015.

[2] 3GPP. Study on new radio access technology: Radio access architecture and interfaces. *3GPP TR 38.801, V.14.0.0*, 2017.

[3] 3GPP. Architecture enhancements for 5G System (5GS) to support network data analytics services. *3GPP TS 23.288 V16.1.0, Rel. 16*, 2019.

[4] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom, J. Medbo, C. Kilinc, and I. Da Silva. Waveform and numerology to support 5g services and requirements. *IEEE Communications Magazine*, 54(11):90–98, November 2016.

[5] 3GPP. NR; Multi-connectivity; Overall description. *3GPP TS 37.340 V15.7.0, Rel. 15*, 2019.

[6] GSMA. Road to 5G: Introduction and Migration. *GSMA Future Networks Programme*, 2018.

[7] 3GPP. Architecture enhancements for control and user plane separation of EPC nodes. *3GPP TS 23.214, V16.0.0, Rel. 16*, 2019.

[8] 3GPP. Interface between the Control Plane and the User Plane Nodes. *3GPP TS 29.244, V16.1.0, Rel. 16*, 2019.

[9] 3GPP. System Architecture for the 5G System (5GS). *3GPP TS 23.501 V16.2.0 , Rel.16*, 2019.

[10] 3GPP. Service requirements for the 5G system. *3GPP TS TS 22.261 V15.2.0, Rel. 15*, 2017.

[11] ETSI. Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services. *ETSI TR 102 643 V1.0.2 (2010-01)*, 2010.

[12] ITU-T. Vocabulary for performance, Quality of Service and Quality of Experience. *ITU-T Recommendation P.10/G.100*, 2017.

[13] Qualinet. White Paper on Definitions of Quality of Experience. *Output from the fifth Qualinet meeting, Novi Sad, March 12*, 2013.

[14] ITU-T. Definitions of terms related to Quality of Service. *ITU-T Recommendation E.800*, 2008.

[15] U. Engelke and H. Zepernick. Perceptual-based Quality Metrics for Image and Video Services: a Survey. In *2007 Next Generation Internet Networks*, pages 190–197, May 2007.

[16] S. Winkler. *Digital Video Quality: Vision Models and Metrics*. Wiley, 2005.

[17] ITU-R. Methodology for the subjective assessment of the quality of Television Picture. *ITU-R Recommendation BT.500-11*, 2002.

[18] ITU-R. Methodology for the subjective assessment of the quality of television pictures. *ITU-R Recommendation BT.500-13*, 2012.

[19] ITU-T. Subjective video quality assessment methods for multimedia applications. *ITU-T Recommendation P.910*, 1999.

[20] ITU-T. Methods for Subjective Determination of Transmission Quality. *ITU-T Recommendation P.800*, 1996.

[21] Ahmet M Eskicioglu and Paul S Fisher. Image quality measures and their performance. *IEEE Transactions on communications*, 43(12):2959–2965, 1995.

[22] ITU-T. The E-model: a computational model for use in transmission planning. *ITU-T Recommendation G.107*, 2015.

[23] ITU-T. P.862 : Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *ITU-T Recommendation P.862*, Approved in February 2001.

[24] ITU-T. P.863 : Perceptual objective listening quality prediction. *ITU-T Recommendation P.863*, Approved in March 2018.

[25] Christos Tselios and George Tsolis. On QoE-awareness through virtualized probes in 5G networks. In *2016 IEEE 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, pages 159–164. IEEE, 2016.

[26] Ying Wang, Peilong Li, Lei Jiao, Zhou Su, Nan Cheng, Xuemin Sherman Shen, and Ping Zhang. A data-driven architecture for personalized QoE management in 5G wireless networks. *IEEE Wireless Communications*, 24(1):102–110, 2016.

[27] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren. Scenarios for 5G mobile and wireless communications: the vision of the METIS project. *IEEE Communications Magazine*, 52(5):26–35, May 2014.

[28] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hofeld, and P. Tran-Gia. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys Tutorials*, 17(1):469–492, Firstquarter 2015.

[29] Sandvine. Global Internet Phenomena. *Sandvine Corp.*, 2015.

[30] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update. *Cisco Corp.*, 2016.

[31] Giorgos Dimopoulos, Ilias Leontiadis, Pere Barlet-Ros, and Konstantina Papagiannaki. Measuring video QoE from encrypted traffic. In *Proceedings of the 2016 Internet Measurement Conference*, pages 513–526. ACM, 2016.

[32] Alessandro Finamore, Marco Mellia, Zafar Gilani, Konstantina Papagiannaki, Vijay Erramilli, and Yan Grunenberger. Is there a case for mobile phone content pre-staging? In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 321–326. ACM, 2013.

[33] How encryption threatens mobile operators, and what they can do about it. `http://goo.gl/fe3xpB`. Accessed: 2017-07-1.

[34] 3GPP. Circuit switched (CS) fallback in Evolved Packet System (EPS). *3GPP TS 23.272, V. 15.0.0*, 2017.

[35] GSMA. VoLTE Service Description and Implementation Guidelines. *GSMA Official Document FCM.01*, 2014.

[36] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between Quality of Experience and Quality of Service. *IEEE Network*, 24(2):36–41, March 2010.

[37] 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*, pages 1–1, 2018.

[38] 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)*, volume 3, pages 2003–2006 Vol.3, June 2004.

[39] Yunhan Jack Jia, Qi Alfred Chen, Zhuoqing Morley Mao, Jie Hui, Kranthi Sontineni, Alex Yoon, Samson Kwong, and Kevin Lau. Performance Characterization and Call Reliability Diagnosis Support for Voice over LTE. In *MobiCom*, pages 452–463. ACM, 2015.

[40] 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*, 66(2):1748–1759, Feb 2017.

[41] 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*, 5:10315–10322, 2017.

[42] 3GPP. Policy and Charging Control Architecture. *3GPP TS 23.203, V. 15.4.0*, 2018.

[43] 3GPP. Mandatory speech codec speech processing functions; AMR speech codec; General description. *3GPP TS 26.071, V.15.0.0*, 2018.

[44] 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*, volume 2, pages 771–774 vol.2, Apr 1997.

[45] 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.

[46] ITU-T. LS on audio issues. *SG 16 Rapporteurs Q7/12*, 2006.

[47] A.D. Clark. Modeling the effects of burst packet loss and recency on subjective voice quality. 2001.

[48] Telchemy's VQmon. `http://www.telchemy.com/vqmon.php`. Accessed: 2018-09-15.

[49] ITU-T. Wideband E-model. *ITU-T Recommendation G.107.1*, 2015.

[50] 3GPP. Codec for Enhanced Voice Services (EVS); General Overview. *3GPP TS 26.441, V. 15.0.0*, 2018.

[51] E. Cipressi and M. L. Merani. A Comparative Study on the Quality of Narrow-Band and Wide-Band AMR VoLTE Calls. In *IEEE 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, pages 1273–1278, June 2019.

[52] Elena Cipressi and Maria Luisa Merani. Effects of packet loss and jitter on volte call quality. *arXiv preprint arXiv:1906.00750*, 2019.

[53] S. Jelassi, G. Rubino, H. Melvin, H. Youssef, and G. Pujolle. Quality of Experience of VoIP Service: A Survey of Assessment Approaches and Open Issues. *IEEE Communications Surveys Tutorials*, 14(2):491–513, Second 2012.

[54] E. Cipressi and M. L. Merani. An Effective Machine Learning Approach to Quality Assessment of VoIP Calls. In *Submission*.

[55] F. Rahdari and M. Eftekhari. Using Bayesian classifiers for estimating quality of VoIP. In *16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*, pages 348–353, May 2012.

[56] L.Sun and E. C. Ifeachor. Perceived speech quality prediction for voice over IP-based networks. In *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002*, volume 4, pages 2573–2577, April 2002.

[57] W. Cherif, A. Ksentini, D. Negru, and M. Sidibe. A_PSQA: PESQ-like non-intrusive tool for QoE prediction in VoIP services. In *2012 IEEE International Conference on Communications (ICC)*, pages 2124–2128, June 2012.

[58] R. S. Alkhawaldeh, S. Khawaldeh, U. Pervaiz, M. Alawida, and H. Alkhawaldeh. NIML: non-intrusive machine learning-based speech quality prediction on VoIP networks. *IET Communications*, 13(16):2609–2616, 2019.

[59] S. Mohamed, F. Cervantes-Perez, and H. Afifi. Integrating networks measurements and speech quality subjective scores for control purposes. In *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213)*, volume 2, pages 641–649, April 2001.

[60] P. Charonyktakis, M. Plakia, I. Tsamardinos, and M. Papadopouli. On User-Centric Modular QoE Prediction for VoIP Based on Machine-Learning Algorithms. *IEEE Transactions on Mobile Computing*, 15(6):1443–1456, June 2016.

[61] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R.* Springer Publishing Company, Incorporated, 2014.

[62] Alan Agresti. *Categorical data analysis.* A Wiley-Interscience publication. Wiley, New York [u.a.], 1990.

[63] Christopher Winship and Robert D. Mare. Regression models with ordinal variables. *American Sociological Review*, 1984.

[64] P. A. Gutirrez, M. Prez-Ortiz, J. Snchez-Monedero, F. Fernndez-Navarro, and C. Hervs-Martnez. Ordinal regression methods: Survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146, Jan 2016.

[65] Hammer. `https://www.empirix.com/products/hammer/`. Accessed: 2019-11-05.

[66] Netem. `https://wiki.linuxfoundation.org/networking/netem`. Accessed: 2019-09-20.

[67] ITU-T. G.1010 : End-user multimedia QoS categories. *ITU-T Recommendation G.1010*, Approved in November 2001.

[68] Wireshark. `https://www.wireshark.org/`. Accessed: 2019-12-30.

[69] Michael Hahsler. Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms - R package. `https://www.rdocumentation.org/packages/dbscan/versions/1.1-4`, 2019. Accessed: 2019-09-08.

[70] Wilhelm Kirch, editor. *Pearson's Correlation Coefficient*. Springer Netherlands, 2008.

[71] Andrew Ng. *Machine Learning Yearning*. Online Draft, 2017.

[72] VASILE Palade. *Class imbalance learning methods for support vector machines*. Wiley Online Library, 2013.