

How to cite this paper:

Ayisat Wuraola Yusuf-Asaju, Zulkhairi Md Dahalin & Azman Ta'a (2018). Framework for modelling mobile network quality of experience through big data analytics approach. *Journal of Information and Communication Technology (JICT)*, 17 (1), 79-113.

FRAMEWORK FOR MODELLING MOBILE NETWORK QUALITY OF EXPERIENCE THROUGH BIG DATA ANALYTICS APPROACH

¹Ayisat Wuraola Yusuf-Asaju, ²Zulkhairi Md Dahalin & ²Azman Ta'a

^{1,2}Department of Computer Science, University of Ilorin, Nigeria

²School of Computing, Universiti Utara Malaysia, Malaysia

ayisatwuraola@gmail.com; zul@uum.edu.my; azman@uum.edu.my

ABSTRACT

The increase in the usage of different mobile internet applications can cause deterioration in the mobile network performance. Such deterioration often declines the performance of the mobile network services that can influence the mobile Internet user's experience, which can make the internet users switch between different mobile network operators to get good user experience. In this case, the success of mobile network operators primarily depends on the ability to ensure good quality of experience (QoE), which is a measure of users' perceived quality of mobile Internet service. Traditionally, QoE is usually examined in laboratory experiments to enable a fixed contextual factor among the participants even though the results derived from these laboratory experiments presented an estimated mean opinion score representing perceived QoE. The use of user experience dataset involving time and location gathered from the mobile network traffic for modelling perceived QoE is still limited in the literature. The mobile Internet user experience dataset involving the time and location constituted in the mobile network can be used by the mobile network operators to make data-driven decisions to deal with disruptions observed in the network performance and provide an optimal solution based on the insights derived from the user experience data. Therefore, this paper proposed a framework for modelling mobile network QoE using the big data analytics approach. The proposed framework describes the process of estimating or predicting perceived QoE based on the datasets obtained or gathered from the mobile network to enable the mobile network operators effectively to manage the network performance and provide the users a satisfactory mobile Internet QoE.

Keywords: Big data analytics, mean opinion score; mobile network operators, telecommunication, users experience.

FRAMEWORK FOR MODELLING MOBILE NETWORK QUALITY OF EXPERIENCE THROUGH BIG DATA ANALYTICS APPROACH

¹Ayisat Wuraola Yusuf-Asaju, ²Zulkhairi Md Dahalin & ²Azman Ta'a

^{1,2}Department of Computer Science, University of Ilorin, Nigeria

²School of Computing, Universiti Utara Malaysia, Malaysia

ayisatwuraola@gmail.com; zul@uum.edu.my; azman@uum.edu.my

ABSTRACT

The increase in the usage of different mobile internet applications can cause deterioration in the mobile network performance. Such deterioration often declines the performance of the mobile network services that can influence the mobile Internet user's experience, which can make the internet users switch between different mobile network operators to get good user experience. In this case, the success of mobile network operators primarily depends on the ability to ensure good quality of experience (QoE), which is a measure of users' perceived quality of mobile Internet service. Traditionally, QoE is usually examined in laboratory experiments to enable a fixed contextual factor among the participants even though the results derived from these laboratory experiments presented an estimated mean opinion score representing perceived QoE. The use of user experience dataset involving time and location gathered from the mobile network traffic for modelling perceived QoE is still limited in the literature. The mobile Internet user experience dataset involving the time and location constituted in the mobile network can be used by the mobile network operators to make data-driven decisions to deal with disruptions observed in the network performance and provide an optimal solution based on the insights derived from the user experience data. Therefore, this paper proposed a framework for modelling mobile network QoE using the big data analytics approach. The proposed framework describes the process of estimating or predicting perceived QoE based on the datasets obtained or gathered from the mobile network to enable the mobile network operators effectively to manage the network performance and provide the users a satisfactory mobile Internet QoE.

Keywords: Big data analytics, mean opinion score; mobile network operators, telecommunication, users experience.

INTRODUCTION

In recent years, immense usage of Internet-based services has been drawn around the evolution of high-speed mobile network located on the Universal Mobile Telecommunication Systems (UMTS), Long Term Evolution (LTE) and other telecommunications (Telecoms) standards. In the same way, the availability of higher data transmission speed (throughput) allows mobile Internet users to go beyond web-surfing by enabling services like file transfer, file download, video streaming and voice-over Internet protocol (VOIP). However, the Network Service Providers (NSPs) or Mobile Network Operators (MNOs) aim to limit the existing data-rate feasible to the users because of the high cost involved in acquiring spectrum (Tsiaras et al., 2014). In most cases, the growth of the Internet subscribers has enhanced competitive advantage and provision of affordable services, at the same time imposing an additional challenge on the MNOs in providing a satisfactory level of network service performance to the mobile Internet users (Ibarrola, Xiao, Liberal, & Ferro, 2011; Shaikh, Fiedler, & Collange, 2010; Tsiaras et al., 2014). Particularly, mobile networks are extremely sensitive to channel availability (such as decreased channel availability) that effectively changes over time because of the local congestion, which often results in compromising the users' session (Goleva, Atamin, Mirtchev, Dimitrova, & Grigorova, 2012). The established instances, an increase in limited data rate and local congestion can severely have a huge influence on the mobile Internet users' experience.

For the MNOs to effectively manage the mobile Internet users' experience, it is imperative to understand that the expectation of the mobile Internet users is based on fulfilled experiences from the network performance (NP), which are generally expected to be stable and less congested. Hence, to facilitate a satisfactory level of users' experience, the MNOs are expected to have detailed knowledge about the traffic characteristics caused by the geographical and dynamic nature of the network traffic (Tsiaras et al., 2014). Having prior knowledge about the users' expectations and network traffic characteristics would assist the MNOs to plan and optimize the NP to understand the geographical and temporal service-related Quality of Experience (QoE) from both the users' and the network's perspective.

QoE is a subjective measure of the perceived quality of mobile Internet services that connect NP, user perception and expectation of the Internet applications

(Chen, Chatzimisios, Dagiuklas, & Atzori, 2016). Considerable effort has been devoted in assessing the QoE of Internet applications through objective and subjective methods over modern fixed and mobile devices (Chen et al., 2016). In most cases, a service-related QoE is often measured through the value of the mean opinion score (MOS) that represents the subjective experience of users for a specific service quality of the network. While several studies have used MOS to measure the QoE of different services such as video streaming (Amour, Souihi, Hoceini, & Mellouk, 2015), VOIP (Charonyktakis, Plakia, Tsamardinos, & Papadopouli, 2016), Skype Voice calls (Spetebroot, Afra, Aguilera, Saucez, & Barakat, 2015) and web-browsing (Balachandran et al., 2014; Rugelj, Volk, Sedlar, Sterle, & Kos, 2014) in laboratory experiments. Limited studies have used large databases obtained from the mobile network traffic constituting the QoE influence factors that usually serve as input for the QoE model (Alreshoodi & Woods, 2013; Balachandran et al., 2014; Tsiaras & Stiller 2014), because mobile network traffic data are not readily available for examination (Tsiaras et al., 2014). In addition, while previous studies presented a specific estimated QoE, usage of diverse possible metrics involving time and location within the mobile network is limited in the literature, as most QoE studies make use of participants in laboratory experiments to aid in the estimation of the QoE measurements (Andrews, Cao, & McGowan, 2006; Tsiaras et al., 2014; Rugelj et al., 2014).

Therefore, to evaluate the users' perceived service-related QoE quantified by MOS, this paper proposed a framework for modelling the mobile network QoE through the big data analytics approach. The proposed framework presented the method involved in analyzing mobile Internet QoE through the data obtained from the mobile network traffic. Utilizing the big data approach would employ the objective measurement gathered from the mobile network traffic for the assessment of the user perceived QoE, by employing different services like file transfer protocol (FTP), Hyper-text transfer protocol (HTTP) and video streaming along with the time and location of the users. Similarly, the usage of big data approach to analyze perceived QoE could assist the MNOs in the allocation of network resources in different geographical areas that might need network optimization to enhance their network service provisioning. The remainder of this article is organized as follows: Section II discusses QoE, perceived QoE influence factors, perceived QoE measurements and perceived QoE modelling. This is followed by Section III which describes big data analytics and the types of big data analytics. Lastly, Section IV presents the proposed framework for modelling the mobile Internet perceived QoE with big data analytics and the methodological instances of the proposed framework.

QUALITY OF EXPERIENCE

The advent of Internet-based services has made QoE gain prominent recognition in the telecoms industry and related research fields. Historically, QoE can be traced back to the operation of NP in mobile network, which is often referred to as Quality of Service (QoS) (Andrews et al., 2006; Chen et al., 2016; Ibarrola et al., 2011). The International Telecommunication Union (ITU), describes QoS as *“totality of characteristics of a telecoms service that bear on its ability to satisfy stated and implied needs of the user of the service”* (ITU-T Recommendation E.800, 2008). Further explanation of QoS by the European Telecommunications Standards Institute (ETSI) supports the view that QoS is the *“collective effect of service performance which determines the degree of satisfaction of a user of the service”* (ESTI, 1994). On the contrary, the Internet Engineering Task Force (IETF) proposes a network-oriented focus by describing QoS as a *“set of service requirements to be met by the network while transporting a flow”* (Crawley, Nair, Rajagopalan, & Sandick, 1998). Evidently, QoS placed more focus on the technical aspects of Internet-based services to enable end-user satisfaction. The technical aspect of the Internet-based services is NP, which constitutes delay, throughput, jitter, loss, and bandwidth of the telecoms network (Chen et al., 2016). Consequently, the wide usage of Internet-based services such as video streaming, VOIP, Skype Voice calls, and web-browsing bring about the assessment of perceived QoS internet services, commonly referred to as QoE (Chen et al., 2016).

Unlike QoS, QoE is a subjective metrics that is concerned with human dimension involving user perception, expectations, experiences of Internet-based applications and NP (Chen et al., 2016). ITU-T Recommendation (2007) defines QoE as the *“overall acceptability of an application or service, as perceived subjectively by the end-user.”* While the definition of QoE provided by ITU focuses on the acceptability of the service, in the Dagstuhl seminar on QoE held in 2009, Fiedler, Kilkki and Reichl (2009) presented an alternative definition that defined QoE as the *“degree of delight of the user of a service, influenced by content, network, device, application, user expectations and goals, and context of use.”* In contrast to the ITU definition which focused on end-to-end system effects and overall acceptability of an application that may be influenced by the user expectations and context (ITU-T Recommendation, 2007), Fiedler et al. (2009) placed emphasis on the quality experience by the user and tacitly considered the network as a QoE influencing factor.

However, recent definition of QoE by Qualinet (Le Callet, Möller, & Perkis, 2012), describes QoE as the *“degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations*

with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state." In contrast to ITU and Fiedler et al's. (2009) QoE definition, by the Qualinet white paper clearly focused on the user by considering the degree of user delight or annoyance with the fulfilment of his or her expectation with time and context. Equally, the description of QoE by the Qualinet white paper indicates that QoE is dependent on QoS and QoS is not enough to understand QoE (Chen et al., 2016; Le Callet et al., 2012). In addition, QoE extends the concept of QoS which is a network-centric approach to a user-centric approach (Raake & Egger, 2014). The user-centric approach of QoE aimed at developing methodological instances for subjective and instrumental quality metrics by considering current and new trends of Internet-based applications along with their application content and interactions (Chen et al., 2016; Möller & Raake, 2014; Raake & Egger, 2014). Generally, users often have predetermined and well-defined expectations that must be met to enable users' satisfaction. In this case, QoE is viewed as a multi-dimensional construct comprising of all the elements influencing users' perception of the network, its performance and how it meets users' expectations (Vuckovic & Stefanovic, 2006). Therefore, QoE is a very vital measure for the MNOs to properly ensure a balance between low quality extremes and over-provisioning of the Internet services. Understanding users' expectations and identifying drivers of users' satisfaction, such as QoE influence factors, are necessary for determining effective perceived QoE measurement and modelling indicators.

PERCEIVED QOE INFLUENCE FACTORS

In the context of telecoms service provision, user experience may be influenced by various factors that impact QoE. QoE influence factors are the characteristics of the services provided by the MNOs to the users. Previous studies have shown that some of the influence factors are clear enough to describe and quantify QoE, while others are situation-dependent, difficult to describe and effective only under certain circumstances (for example in combination with or without other influence factors (Reiter et al., 2014). The Qualinet white paper defines QoE influence factors as *"any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the QoE for the user"* (Le Callet et al., 2012). In this case, the influence factors are the independent variables while the resulting QoE as perceived by the user is the dependent variable (Reiter et al., 2014). Oftentimes, a certain set of influence factors may be noticeable by the users in terms of the impact on users' perceived QoE. In other words, users may not necessarily be aware of the underlying influence factors, but to some extent

the users can describe what they like or dislike about their perceived QoE. The QoE influence factors can be classified into different dimensions as depicted in Table 1 below.

Table 1

Dimensions of QoE Influence Factors

Authors	Dimensions	Components
Barakovic, Barakovic and Bajric (2010)	Technology performance	Application/service, server, network and device.
	Usability	Behavioural usability, ease of use, device features, emotions and feelings.
	Expectations	Application type, usage history, gender, brand and personality.
	context	Environment, personal, social context, technological context and cultural context.
	Subjective evaluation	Service, network and device.
DeMoor et al. (2010)	QoS parameters	Delay, jitter, loss, throughput and bandwidth.
	Context, Prior experiences, Expectations	Place of use and historical experience.
Stankiewicz and Jajszczyk (2011)	User Factors	Personalisation and emotions.
	QoS factors, Grade of Service (GoS), Quality of Resilience (QoR)	Terminals, type of content, application specific features.

(Continued)

Authors	Dimensions	Components
Stankiewicz and Jajszczyk (2011)	Emotions, occupation, education level and age.	Customer profiles, environmental, psychological and sociological aspects.
	Pricing policies	Prepaid or Postpaid.
Skorin-Kapov and Varela (2012)	Application	Application configuration-related factors.

(continued)

Authors	Dimensions	Components
	Resource space	Delay, jitter, loss, throughput and system-related factors).
	Context	Customer location, time, and application-related factors.
	User space	Demographics, customer preferences, requirements, expectations, prior knowledge, behaviour and motivations.
Barakovic and Skorin-Kapov (2013) Le Callet et al. (2012)	Human factors	Age, education background, emotions, gender and user visual aid.
	System factors	Bandwidth, delay, loss, throughput, security, display size and resolution.
	Context factor	Location, movement, time of day, costs, subscription type and privacy.

However, evidence has shown that all the QoE factors discussed in prior studies cannot be addressed in a single study to analyze perceived QoE (Barakovic & Skorin-Kapov, 2015). Therefore, recent studies supported three dimensions (human, system, and context) and justified that the three dimensions are essential for modelling QoE as perceived by the customers (Barakovic & Skorin-Kapov, 2015; ITU-T Recommendation P.10/G.100, 2016; Reichl et al., 2015).

The human influence factor is a dimension of the QoE influence factor that describes any characteristics of human users such as the demographic, socio-economic background, physical and mental constitution, or emotional state (Le Callet et al., 2012; Reiter et al., 2014). Previous theoretical and conceptual studies have highlighted the importance of human influence factors and the possible effects on QoE (Geerts et al., 2010; Laghari, Crespi, & Connelly, 2012; Reiter et al., 2014). Additionally, to a certain extent, some studies have investigated the impact of certain human factors on perceived QoE (Quintero & Raake, 2011; Wechsung, Schulz, Engelbrecht, Niemann, & Moller, 2011). Equally, human influence factors have been taken to a limited extent in most empirical studies, due to the difficulties involved in assessing some of the human influence factors (Reiter et al., 2014; Sackl, Masuch, Egger, & Schatz, 2012). Some examples of human influence factors are gender, age, background,

emotion and education (Le Callet et al., 2012; Reiter et al., 2014). However, inherent complexity and lack of empirical evidence has left an impact of the human influence on perceived QoE to be poorly understood (Reiter et al., 2014).

Another dimension of the QoE influence factor is the system influence factor constituting the properties and characteristics that determine the technically produced quality of an application or service (Le Callet et al., 2012). The system influence factor comprises of content, network, and device-related factors. Content-related factors includes graphical design elements, sematic content, video spatial and temporal resolution, depending on the kind of application or services being used (Chen et al., 2016). The network-related influence factor is made up of the QoS parameters (such as throughput, delay, jitter and loss) and security (Le Callet et al., 2012), while the device-related influence factor specifies the characteristics and capabilities of the devices located at the end points of the communication path (Chen et al., 2016).

The last dimension of the QoE influence factor is the context influence factor that deals with any situational property to describe the users' environment (Le Callet et al., 2012). Previous studies usually combined context factors with human and system factors without any specific structure or categorization (Reiter et al., 2014). However, in the mobile network scenario, context factors were broken down into physical, temporal, social, economic, task and technical components (Jumisko-Pyykko, Satu, & Vainio, 2010) The physical components of the context influence factor describe the characteristics of location and space along with the movements within and transitions between locations (Reiter et al., 2014). Generally, user preferences can vary in different contexts such as location, time movement and mobility (Jumisko-Pyykko, Satu, & Vainio, 2010; Reiter et al., 2014). Therefore, the physical components of the context influence factor are essential for analyzing the perceived QoE of mobile Internet users. Another component of context influence factor is temporal component, which describes the past and future situations involving the time of the day, month, and year (Jumisko-Pyykko, Satu, & Vainio). The social component is another type of the context influence factor that defines the inter-personal relation existing during the experiences observed through the mobile network (Reiter et al., 2014). Some examples of the social component are cultural, educational and professional levels (Reiter et al., 2014). The economic component is also an important component of the context influence factor that comprises of costs, subscription type or brand of the application or system used by the users (Reiter et al., 2014). Task is another type of context influence factor that determines the nature of the experience depending on the user situation (Reiter et al., 2014). Some authors concluded that an additional

task does not have influence on the perceived quality, independently of the difficulty of the task (Sackl, Seufert, & Hoßfeld, 2013). But the authors' conclusion does not limit the importance of the task component on the context influence factors because the application used by the user may have a huge impact on the perceived QoE of the user. The last component of the context influence factor is the technical component that describes the relationship between the system and the devices (Reiter et al., 2014). Some examples of the technical components are applications and network components.

Generally, the most studied QoE influence factor is the system influence factor constituting the QoS parameters (throughput, loss, bandwidth, delay, and jitter) and the technical component that is a subset of the context influence factors (Alreshoodi & Woods, 2013). While there exist many studies that examined throughput measurement for wireless applications for web traffic (Barakovic & Skorin-Kapov, 2013; Rugelj et al., 2014; Singh et al., 2013), few studies used the user experience measurements obtained from the mobile network traffic to model perceived QoE, as most studies gathered basic network performance measurement data in laboratory experiments through the desktop applications (Rugelj et al., 2014; Singh et al., 2013). Gathering measurement data from the desktop application in laboratory experiments limits the use of physical (location, time movement and mobility), temporal components (the past and future situations involving the time of the day, month, and year) and economic components constituted in the context influence factors (Barakovic & Skorin-Kapov, 2013; Tsiaras et al., 2014). Therefore, it crucial to examine specific service-related throughput in mobile network traffic in relation to expectation, mobility, (location and time) and different services like FTP, HTTP, and video streaming. On this basis, it is important to gather user experience measurement from the mobile network traffic to analyze the perceived QoE from both the network and users' perspectives.

PERCEIVED QOE MEASUREMENTS

Based on the classification of the QoE influence factors discussed above, it should be noted that measuring and analyzing perceived QoE could be challenging due to the complexities involved in capturing the user's experience metrics (K. Laghari, Issa, Speranza, & Falk, 2012). Perceived QoE is an assessment of users' expectations, perception, cognition and satisfaction with respect to a specific application or service (K. Laghari et al., 2012). In most cases, perceived QoE assessment is presented through MOS, which is a five-point Likert scale (5=Excellent, 4=Good, 3=Fair, 2=Poor, and 1=Bad) metrics used to quantify perceived QoE (Raja & Flanagan, 2008; Streijl, Winkler, &

Hands, 2016) of Internet-based applications. MOS is an average score across subjects that has been widely used in numerous applications for both subjective and objective measurements in laboratory testing and in-service monitoring.

Subjective and objective measurements are two types of perceived QoE measurements. Subjective measurement is commonly based on controlled real-life experiments that involve users' participants who directly evaluate their experience of an application or service (Tsolkas, Liotou, Passas, & Merakos, 2016). The users involved in subjective measurement can both be in active or passive form and judge their perceived experience. Equally, the users in the experiment can score their perceived quality using an absolute rating scale as well as compare sequential service-related experience. The results of the subjective measurement are often based on user opinions, previous experience, expectation, user perception, judgement, description capabilities, effectiveness, efficacy and overall capabilities of using a service (Tsolkas et al., 2016). Previous studies termed subjective measurement as a reliable measurement because they incorporate the conscious and unconscious aspects of the users' quality of evaluation aspects that may otherwise not be captured (Barakovic & Skorin-Kapov, 2013; Rugelj et al., 2014; Shaikh et al., 2010; Singh et al., 2013; Tsolkas et al., 2016). In addition, subjective measurements are considered reliable if the process is designed carefully and unbiased (Tsolkas et al., 2016). However, one major drawback is that the subjective measurements are valuable only for the laboratory testing of some services and not visible in real-time QoE evaluation and support (Alreshoodi & Woods, 2013; Andrews et al., 2006; Barakovic & Skorin-Kapov, 2013; DeMoor et al., 2010; Shaikh et al., 2010; Singh et al., 2013; Tsolkas et al., 2016). Other drawbacks of the subjective measurements are time-consuming, costly, and are not reproducible on demand (Tsolkas et al., 2016). Thus, subjective measurement may not be efficient for in-service quality monitoring (Tsolkas et al., 2016). One way to overcome these drawbacks is to conduct real-service QoE evaluation, where users' experience can be captured and evaluated in real-time (Tsolkas et al., 2016). As a result, the drawback gave rise to an objective measurement that can measure or predict the quality perceived by the users without the users' intervention.

In contrast to the subjective measurement is the objective measurement, which aims to predict human behavior using a mathematical formula/model rather than getting direct feedback from the end users (Shaikh et al., 2010; Singh et al., 2013). Objective measurement is preferred by most authors because of its ability to be implemented and be embedded into the network using software applications (Falk & Chan, 2006) and the capability to allow researchers to model the relationships that exist within the user's experience metrics to

determine the MOS or users' perceived QoE (Sharma, Meredith, Lainez, & Barreda, 2014). An example of the objective model is the parametric model that uses network planning parameters and measures the values of specific network metrics. The parametric model based its estimations on parameter metrics collected at runtime from network process and control protocols (Tsolkas et al., 2016). Another example of the objective measurement is the use of hybrid methods, based on employing machine-learning algorithm on the user's experience metrics gathered from the network. The user experience metrics do represent the QoE influence factors and are used as input to train the perceived QoE model. In other words, the model obtained from the hybrid methods maps the QoE influence factors to MOS values and further use the model for real-time quality prediction. Presently, most objective models account for the user factors in terms of their inherent characteristics but the context and content of the services are only considered at a limited extent (Rugelj et al., 2014; Shaikh et al., 2010; Tsolkas et al., 2016). To enable the consideration of context and content of Internet service-related applications in the mobile network, there is a need to design more accurate objective estimation models that adopt the use of both hybrid and parametric methods to enable an indirect and user-transparent perceived QoE model (Liotou, Tsolkas, Passas, & Merakos, 2015; Tsolkas et al., 2016) to assist the MNOs overcome the challenges associated with QoE management in mobile networks.

PERCEIVED QOE MODELLING

Perceived QoE modelling is used to quantify the QoE influence factors by defining a correlation or prediction model that estimates the MOS. MOS is used as the linkage between the subjective test and the objective modelling along with other quantitative information. The usage of MOS enables the overall measurement of the network from the users' perspective. Though the factors influencing QoE are specific to certain applications, the factors that influence video applications may be different from web-browsing applications. In most cases, the QoE influence factors are considered as the predictors while the predicted outcome is the perceived QoE/MOS, so it is imperative to find the correlation between the influence factors and the perceived QoE. Hence, accurate service-related applications measurement and monitoring at different system nodes will enable the MNOs to achieve maximum user perceived QoE.

Several studies have investigated the correlation between the QoE influence factors to determine the estimated MOS of the users. The study of Fiedler, Hossfeld and Tran-Gia (2010) indicates the QoS parameters (such as loss delay, jitter and throughput) in the system QoE influence factors can translate

into user experience instances like excessive waiting time (longer time taken by users to access the internet applications). Equally, another study points out that the response time is very essential when relating these QoS parameter to the perceived experience (Shaikh et al., 2010). The challenges often observed in most studies are how to link or map the quantitative metrics of the QoS parameters with the perceptual quality of the customers (Reichl, Egger, Schatz, & D’Alconzo, 2010). Therefore, a mathematical interdependency was developed using logarithmic relationship and exponential interdependency between the QoS parameters and QoE. The study argued this based on the Weber-Fechner law that describes the relationship between QoS parameters and other QoE influence factors as the stimulus-perception of human sensory (Reichl et al., 2010). The law states that “*just noticeable difference*” between two levels of stimulus is proportional to the magnitude of the stimuli (Barakovic & Skopin-Kapov, 2013; Reichl et al., 2010; P. Reichl et al., 2011). The law further explains that the perception dP , to be directly proportional to the relative change dS / S of the physical stimuli of size S as presented in Equations 1 and 2 (Reichl et al., 2010; P. Reichl et al., 2011).

$$dP = k \cdot \frac{dS}{S} \quad \text{Equation 1}$$

Integrating Equation 1, would result in $P = k \cdot \ln \frac{S}{S_0}$ Equation 2

where P describes the magnitude of perception and the constant integration S_0 is interpreted as the stimulus threshold. This law is valid for a wide range of scenerios like hearing, time perception and even numerical cognition. In the case of QoS parameters and QoE, Reichl et al. (2010) mentioned the the QoS parameter (such bit rate) to represent stimulus S , and QoE as the perception P . Hence, the propotionality can be expressed as $dQoS \propto QoS \cdot dQoE$.

On the other hand, the exponential interdependency, in contrast to the Weber-Fechner law, is based on the IQX hypothesis which describes QoE as the level of perception and QoS parameters as the level of disturbance (Fiedler et al., 2010). The IQX hypothesis describes $QoE = \phi(I_1, I_2, \dots, I_n)$ as a function of n influence factors I_j (Fiedler, Hossfeld, & Tran-Gia, 2010). For instance, using a single QoS parameter such as throughput, that is $I = QoS$, then, the fundamental relationship would be $QoE = f(QoS)$. This means that the subjective sensibility of QoE would be more pronounced and the higher than the experienced quality is observed (Fiedler et al., 2010). For instance, if the QoE is very high, a little deterioration will strongly decrease QoE. The overall analysis means that the change in QoE depends on the present level of QoE, given the same amount of change of QoS (Fiedler et al., 2010). This is expressed as: $\frac{\partial QoE}{\partial QoS} \sim -(QoE - \gamma)$. This equation is an exponential function and it expresses the fundamental

relation of the IQX hypothesis. Because Fiedler et al’s (2010) study on IQX hypothesis lacks perceivable stimulus translation, the IQX hypothesis was enhanced by translating it into a perceptual change for a given fixed change of the stimuli proportional to the current level of perception (Reichl, Egger, Schatz, & D’Alconzo, 2010). This relates to changes in QoE with respect to QoS to the current level of QoE expressed as: $QOE = \alpha \exp(-\beta * QoS) + \gamma$, where α, β and γ are positive parameters (Alreshoodi & Woods, 2013). However, the drawback of the IQX hypothesis is that it only considers the use of one QoS parameter at a time and only focuses on the quality o deterioration parameters (P. Reichl et al., 2011).

A large and growing body of literature has adopted the approach of the IQX hypothesis for modelling perceived through the machine-learning algorithms (Amour et al., 2015; S. Aroussi & Mellouk, 2014; Spetebroot et al., 2015). Machine-learning algorithms is a technique that designs and develops algorithms capable of building a reality model from the data, either by improving the existing model or building a new model (S. Aroussi & Mellouk, 2014). Machine-learning algorithms aimed at correlating QoE influence factors through prediction, which focus on some known properties or acquired from an observation that reflects both the network and customer’s perception (S. Aroussi & Mellouk, 2014). Decision Tree, Random forest, Support vector machine, K-nearest and artificial neural network are the most commonly used machine learning algorithms for the modelling of perceived QoE (Amour et al., 2015; S. Aroussi & Mellouk, 2014; Aroussi & Mellouk, 2016; Spetebroot et al., 2015). Table 2 depicts previous studies that have used machine-learning for modelling perceived QoE.

Table 2

Modelling Perceived QoE with Machine-learning Algorithms

Authors	Dataset/Scenerio	Application/ Service type	Machine-learning algorithms
Anchuen, Uthansakul, and Uthansakul, (2016)	Network tool/ Experiment	Smartphone	Neural network
Li et al. (2016)	Participant data/ Experiment	Over-the-top video	Decision Tree

(continued)

Authors	Dataset/Scenerio	Application/ Service type	Machine-learning algorithms
Charonyktakis et al. (2016)	Test bed experiment	VOIP	Decision Tree, Gaussian naïve bayes, Artificial neural network and support vector machine
Aroussi and Mellouk (2016)	Testbed Experiment	Video on Demand (VoD)	Artificial neural network, K-nearest, Support vector machine, Decision Tree, Naïve bayes and Random forest
Amour et al. (2015)	Participant data/ Laboratory experiment	Video	Naïve bayes, Decision Tree, Random forest, Support vector machine and Neural network
Spetebroot et al. (2015)	Testbed experiment	Skype voice calls	Decision Tree, Rule induction, Logistic regression, Support vector machine, Neural network, Lazy learners and Ensemble method.
Balachandran et al. (2014)	Mobile websites data / Cellular network	Mobile web browsing	Text classification (Decision Tree and Linear regression)
Rugelj et al. (2014)	Participant data/ Laboratory experiment	Web-browsing	Exponential Regression and Hidden Markov Model.
Mushtaq, Augustin and Mellouk (2012)	Participant data/ experiment	Testbed video	Naïve bayes, Decision Tree, Random forest, Support vector machine, K-nearest and Neural network
Calyam et al. (2012)	Testbed experiment	IPTV	Neural network
Hoßfeld et al. (2011)	Participant data/ Laboratory experiment	Web-browsing	Exponential regression, Support vector machine and Hidden memory markov models

(continued)

Authors	Dataset/Scenerio	Application/ Service type	Machine-learning algorithms
Machado et al. (2011)	Testbed experiment	Multimedia streaming	Artificial neural network
Du, Guo, Liu and Liu (2009)	Simulation Test Experiment	Video	Neural network
Menkovski, Cuadez-Sanchez, Oredope and Liotta (2009)	Testbed experiment	Video	Support vector machine and Decision Tree

As indicated in Table 2, the prevailing method of modelling perceived QoE is through the testbed experiment often conducted in a laboratory. The testbed experiment can be in different forms depending on the method adopted by the researcher. A testbed experiment was conducted in K. Laghari et al’s (2012) study by setting up a private local area network (LAN) using two laptops connected to a gateway through a switch. The testbed was used to emulate the wireless environment in order to analyze the effects of varying network conditions on video streaming QoE. Specifically the study considered packet loss (PLR) as a QoS parameter involving packet re-order (PRR) and video bit rate (VBR). A user experiment was conducted with 33 subjects (25 males and 8 females). They were provided with questionnaires and asked to provide their profile information and feedback about the perceived video quality (PVQ) using a 5-point scale, where label ‘1’ corresponded to “Worse/Strongly dissatisfied” and label ‘5’ to “Excellent/Strongly satisfied”.

Another testbed experiment was conducted in a controlled environment with sufficient light and air to produce consistent and reproducible results. The interactive Graphical User Interface (GUI) was used in the study and subjective scores were collected from the GUI and stored in the database (Battisti, Carli, & Paudyal, 2014). A similar testbed experiment was conducted through the Distributed Passive Measurement Infrastructure (DPMI) constituting a server, a client, the Linux Traffic Controller (TC) shaper, two measurement points (M2 and M3), a measurement area controller and the consumer station for data (Shaikh et al., 2010). Other testbed experiment studies often involved volunteered participants of different age groups to collect data in a controlled environment with a high level of control to enable the estimation/prediction of the perceived QoE for different internet applications (the likes of web browsing, video and VOIP applications (Alreshoodi & Woods, 2013; Aroussi

& Mellouk, 2016; Calyam et al., 2012; Charonyktakis et al., 2016; Calyam et al., 2012; DeMoor et al., 2010; Fiedler et al. 2010; Geerts et al., 2010; Li et al., 2016; Menkovski et al., 2009; Rugelj et al., 2014; Spetebroot et al., 2015)

Similarly, as seen in Table 2, most of the studies focused on a specific application or service, because the authors tended to make the contextual factors as fixed as possible for a certain QoS parameter which is a variable of the system QoE influence factor. To overcome this drawback, a deterministic mathematical model (DQX) was proposed by Tsiaras et al. (2014) to measure the impact of QoS parameters and other influence factors on QoE. The study defined service-specific QoS values through the DQX model for quantifying the QoS parameters to QoE. The DQX model overcame the drawback of the IQX hypothesis by considering multiple QoS parameters as input. In addition, the DQX model examined the positive and negative impacts of QoS on QoE rather than just a deteriorating effect as in the case of the IQX hypothesis. The DQX model allows flexibility of the QoS parameters by using the concepts of the expected variable value and expected MOS. This simply means that a certain level of QoE can be maintained even if one variable changes. The formalization of the QoE is given by $QoE := f(User, Service, Variable)$ (Tsiaras et al., 2014). The overall analysis of the DQX model enables the use of multiple and diverse parameters and explains how the parameters can affect the perceived QoE positively or negatively in a specific situation (Tsiaras et al., 2014). A broader perspective of the DQX model was applied on the Voice-over Internet protocol-based (VOIP) using an experimental set up to capture all the end-users of QoE data in a VOIP services (C. Tsiaras, Rösch, & Stiller, 2015). The data was used to define all the necessary parameters such as latency, jitter, packet loss and bandwidth in VOIP scenarios. The results showed that the DQX model produced promising results, especially on the measurements with the mixed QoS parameters. The study revealed that the DQX model was precise, highly adaptable, and concluded that the DQX model was a powerful and useful tool for MNOs to predict and improve their services in relation to perceived QoE. Evidently, the idea of the DQX model supports that QoE perceived can be optimized to determine the actual perceived QoE, because it supports the use of multiple QoS parameters along with other QoE influence factors with regards to the minimum, maximum, expected variable values and variable weights to enable the modelling of the perceived QoE (Aroussi & Mellouk, 2016; Tsiaras et al., 2014). Despite the importance of the DQX model to determine the expected MOS which represents the perceived QoE, it has not been applied in the mobile environment that comprises of a large scale scenario (C. Tsiaras et al., 2015; Tsiaras & Stiller 2014).

However, evidence has shown experimentally that there is a need to quantify QoE of different mobile internet applications in relation to time and location

within a mobile network (Barakovic & Skorin-Kapov, 2015; Reichl et al., 2015; Tsiaras et al., 2014). To quantify perceived QoE of mobile applications, Tsiaras et al. (2014) used an android application to gather the QoE representation in MOS values because the android application was already designed to evaluate multiple performance measures. The study only focused on gathering data from an android application by instructing the clients/users to create traffic that contains a GET request for a web page from Wikipedia (Tsiaras et al., 2014). In this case, the data gathered could be restricted to a certain set of clients in a certain location, because the clients must be given instructions on a specific website that the measurement test data needs to be collected. Considering the increase in the volume of broadband data traffic of the mobile network caused by the diverse and large amount of mobile internet users, recent literature suggests the need for an advanced QoE management scheme and optimization algorithms for both the wireless and mobile systems (Aroussi & Mellouk, 2016; Rugelj et al., 2014). The advanced QoE management scheme may involve the process of gathering large user experience in relation to user behavior from the mobile network traffic (Reichl et al., 2015). Such large user experience data is fundamentally a big data problem, and requires some big data analytics for such data to be effective and analyzed (Spiess, T'Joens, Dragnea, Spencer, & Philippart, 2014). Therefore, this study suggests the use of the minimum, maximum, expected variable values and variable weights stated in the DQX model for an analytical and large-scale scenario to determine the correlation and mappings of the QoE influence factors to enable the estimation of the perceived QoE of the mobile internet users and to enable maximization of QoE, to determine the actual customer satisfaction in relation to the customers, expectation as stated in the service level agreement (SLA).

BIG DATA ANALYTICS

Big data is a collection of large amount of data that has the ability of changing rapidly over a particular period (Spiess et al., 2014). In recent times, most organizations especially the telecoms organizations are much more interested in data-driven decisions due to the large and diverse dataset generated within the mobile network traffic. The data-driven decisions are of great importance to the MNOs to enable them to deal with disruptions as observed in the NP and provide an optimal solution based on the insights (information and knowledge) derived from the data (Spiess et al., 2014). Big data constitutes five major characteristics such as volume, velocity, variety, value, and veracity. Volume constitutes the mass and quantity of the data. Velocity involves the speed of data creation that is, how quick the data is generated and processed to meet the present network demand and prepare for future challenges. Variety constitutes

different kinds of data, most especially the classes of data generated in the same network traffic. Veracity depicts the accuracy, correctness, quality of data sources and the uncertainties observed in the dataset. Value characterizes the type of insight that can be extracted from the supposed big data. ITU (2014) affirms that the data generated in the mobile network traffic constitutes all these five characteristics, thus the large dataset collected from the network traffic can be used for perceived QoE modelling, estimations, and monitoring in a diverse heterogeneous environment that is very essential for network optimization (Zheng et al., 2016). In addition, employing the usage of big data can assist the MNOs to prevent future occurrences of network problems, proper allocation of infrastructural resources in different geographical area, and allow the selections of accurate key indicators to measure and improve user experience (Zheng et al., 2016).

TYPES OF BIG DATA ANALYTICS

Analytics is a technique used in analyzing the large dataset (big data) generated from the mobile network traffic. There are three different types of big data analytics; descriptive, predictive, and prescriptive analytics. Descriptive analytics is a process of using exploratory analysis comprising of statistical techniques such as central tendency (mean, median, and mode), measures of dispersion (standard deviation), charts, graphs and frequency distribution to aid the understanding and visualization of the big datasets. Descriptive analytics using exploratory analysis allows the grouping of data through the distribution of values and interrelationships within the dataset to determine the presence of extreme values present in the dataset and the discovery of high-level patterns in data that facilitates the understanding of the dataset easily (Kotu & Deshpande, 2015).

Predictive analytics takes a step further than descriptive analytics when data is used to seek the future state of business performance. Predictive analytics originated from artificial intelligence, statistics, machine-learning, and data-mining techniques. Predictive analytics aims at predicting the probability of the future occurrence of patterns or trends in data. Moreover, predictive analytics is sometimes referred to as one-click data mining, because it simplifies and automates the data-mining process to discover the factors leading to specific outcomes, as well as predict the likely outcomes with a degree of confidence in the predictions (Deka, 2014). One of the advantages of predictive analytics is its ability to predict network outages through the analysis of customer complaints and network data. Another advantage is the prediction of valuable customer segment that can be used for customer retention campaigns. These

advantages of predictive analytics can assist the MNOs to identify the root causes of network failures and direct retention campaigns to a focused group to generate average revenue per user (ARPU) and increase the spending of loyal customers (Spiess et al., 2014).

Prescriptive analytics is the last stage of big data analytics technique. It is sometimes referred to as optimization analytics, since organizations can use it to optimize their scheduling, production inventory and supply chain design. Prescriptive analytics suggest decision options with their implications. An example in mobile telecoms industry is the allocation of infrastructural resources to locations which would enable the MNOs to operate their networks more efficiently (Zheng et al., 2016). Prescriptive analytics adopts the use of mathematical programming, heuristic search and simulation modelling to identify the optimum actions to be taken by MNOs to improve their NP. Collectively, the use of big data analytics to manage user experience in the mobile network can assist the MNOs to have an adequate insight on the most important user experience measurements (such as total throughput, download transfer time and connection duration) that can impact the perceived users' experience.

In most cases, the use of big data analytics often aids the processing and understanding of the large diverse dataset. Equally, employing machine-learning and data mining algorithms usually aid the discovery of knowledge insights about the large datasets. In addition, processing and analyzing big data provides an automatic and speedy solution in dealing with real-life problems, thereby facilitating human understanding of the medium of data analysis outcome through the process of visual representation. Therefore, employing big data analysis for the QoE management scheme will permit an early insight about mobile internet customer behavior such that timely actions can be taken early to improve user experience. At the same time, the prediction of perceived QoE through previous users' behavior would assist the MNOs to provide an optimal NP and rectify the occurrence of any outage in the service utilities before the customers would experience it.

PROPOSED FRAMEWORK FOR MODELLING MOBILE NETWORK PERCEIVED QOE USING BIG DATA ANALYTICS APPROACH

Modelling of perceived QoE is concerned with the process of predicting the perceived QoE of users through an abstract representation of data and

its relationship from the information collected from the users, network, or both users' networks, considering the drawbacks of context and content of the service observed in the objective measurement of perceived QoE, limited use of large and diverse dataset generated in mobile networks stated in previous sections (Alreshoodi & Woods, 2013; Machado et al., 2011; Reichl et al., 2015). In addition, based on the challenges faced by the MNOs in analyzing the vast amount of data constituted in the mobile network (Diaz-Aviles et al., 2015; Spiess et al., 2014), this section proposed a framework that enables modelling of perceived QoE in the mobile network through the big data analytics approach, because the MNOs could use big data analytics to have a clear and current understanding of the users' experience to enable them measure and model perceived QoE of the mobile Internet users (Diaz-Aviles et al., 2015). The proposed framework describes the process of gathering data from the mobile network traffic and the three processes of big data analytics in the real time measurement platform as depicted in Figure 1.

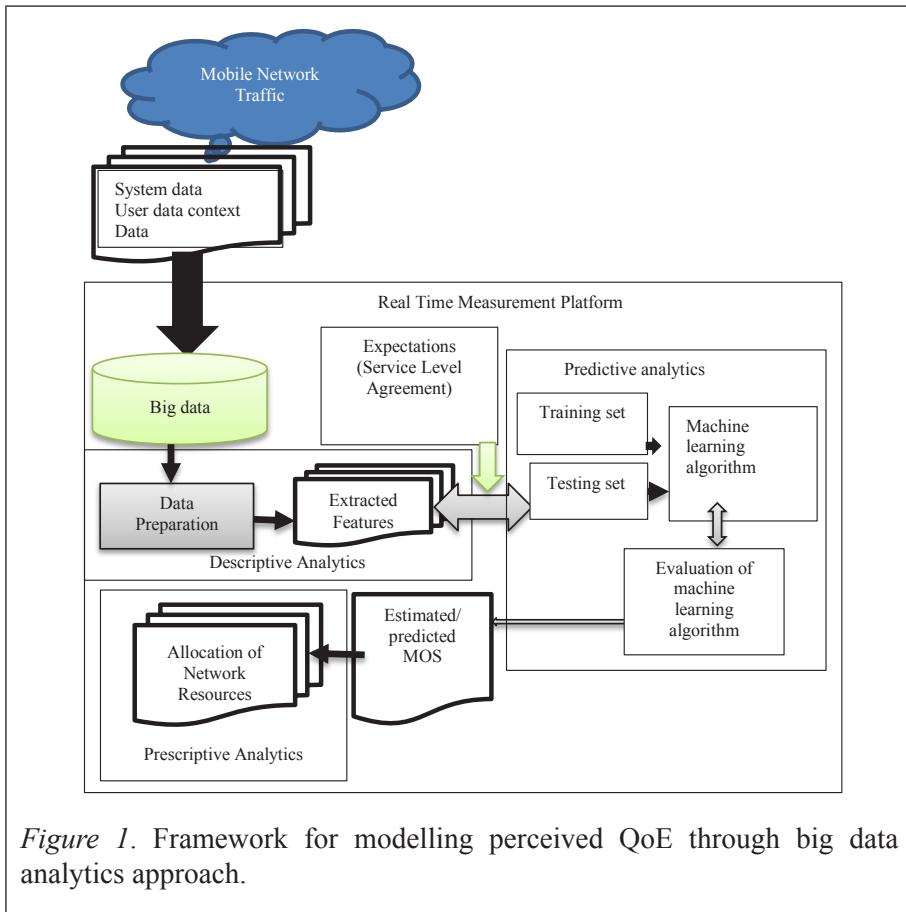


Figure 1. Framework for modelling perceived QoE through big data analytics approach.

This study argued based on the approach of using large datasets obtained from the mobile network for the modelling of Internet service-related applications perceived QoE. To this end, this study supported the view that the mobile network is made up of large diverse key quality indicators (KQI) and key performance indicators (KPI) datasets consisting of many files from a vast number of cells (Yang, Liu, Sun, Yang, & Chen, 2016). The KQI is a quantitative measures of key system elements performance that is relevant to customer's needs and expectations such as the translation of a rate to frequency in a tangible perception from the customer's view (ETSI, 2014). The KPI emanates from the definition of the key parameters measurement of input and output NP (ESTI, 2010). In short, KQI and KPI are often used to indicate the service resource performance of the network. These KPI and KQI constitute the perceived QoE influence factors that can be used for measuring and analyzing the Internet service-related application perceived QoE. Each of the files contain in the cell-level of KQI and KPI, values of all users over a period of time for instance, a week, months or even years. Values attached to this aggregated or averaged KQI and KPI are generated over a predefined time interval of two, five or ten minutes (Yang et al., 2016). Some examples of these KPI and KQI are the download bit rates, upload bit rates, latency, time, date, longitude, and latitude (Anchuen et al., 2016).

The KPI and KQI are often extracted through the pre-processing of the raw dataset gathered from various network elements and probes (Deka, 2014). KPI and KQI are extracted from the pre-processing process because the dataset contained in mobile network traffic is assumed to be inconsistent and dirty due to the voluminous nature of the dataset (Mohanty, Jagadeesh, & Srivatsa, 2013; Tsai, Lai, Chao, & Vasilakos, 2015). In addition, big data constituting the KPI and KQI are often available in an unstructured form that may not be suitable for the modelling of perceived QoE. The data pre-processing phase of the big data analytics will ensure reliability, completeness, randomness, and consistency of the dataset to make it suitable for the perceived QoE modelling phase (Mohanty et al., 2013). In most cases, reliability of the dataset will ensure the represented dataset is accurate enough to suit the perceived QoE modelling phase. The randomness of the datasets describes the statistical characteristics of the complete datasets, which is very essential for exploratory data analysis and visualization of the dataset. Then the consistency of the data will ensure the dataset produce the same result within an acceptable error margin when a different random sample analysis is conducted (Mohanty et al., 2013; Tsai et al., 2015). In this case, usage of exploratory data analysis and traditional data pre-processing methods such as data cleaning, data integration, data reduction and data transformation are commonly used in the data-mining technique; feature selection and extraction will effectively assist

the big analytics methodology to aid the process of modelling perceived QoE of the mobile Internet users (Atzmueller, Schmidt, & Hollender, 2016; Tsai et al., 2015). As a result, the proposed framework adopts the use of big data obtained from the mobile network traffic consisting of various KPI and KQI, which represent the perceived QoE influence factors as the core foundation for modelling perceived QoE of mobile Internet service-related applications.

It is worth mentioning that the use of expectation in the form of service level agreement (SLA) is an important parameter for modelling perceived QoE, but the use of SLA is still limited in the literature (Tsiaras & Stiller 2014). The common method for using expectation in modelling perceived QoE is by asking the users what is expected from the MNOs through the process of a survey (that is, subjective method (Rugelj et al., 2014), because most studies assumed that user expectation grows as network and applications continually developed (Rugelj et al., 2014). But considering the time consuming and expensive nature of the subjective method used in gathering individual user expectations (Falk & Chan, 2006; Shaikh et al., 2010; Singh et al., 2013), subjective measurement may not be suitable in large-scale settings. Moreover, the subjective method lacks repeatability and is not effective in real-time scenarios (Alreshoodi & Woods, 2013; Barakovic & Skorin-Kapov, 2013). However, in the case of the objective method where the users' experience would be captured and evaluated in real-time without direct feedback from the users' it is vital to use SLA along with other QoE influence factors to estimate the users perceived QoE. SLA is the agreement between the customer and the MNOs on service characteristics, such as service level objectives, service monitoring components and financial compensation components (Gozdecki, Jajszczyk, & Stankiewicz, 2003). The telecoms regulators often use SLA to assess the whether the services provided by the MNOs comply with the criteria stated in the agreement. Therefore, SLA is incorporated in the proposed framework as suggested in the recent studies (Tsiaras et al., 2014; Tsiaras & Stiller 2014). Employing SLA in the proposed framework for modelling perceived QoE would aid the MNOs to determine when one or more variables do not meet the expected level stated in the SLA and how exactly the variables involved impact user experience (Tsiaras & Stiller 2014). Overall, using SLA as user expectation in modelling perceived QoE would aid the process of determining the expected MOS, based on the maximum and minimum values stated in the SLA.

In addition, the proposed framework incorporated the three types of big data analytics (descriptive, predictive, and prescriptive) methods discussed in previous studies (Spiess et al., 2014; Zheng et al., 2016). Following the advantages of the big data analytics discussed in prior studies (ITU, 2014;

Spieß et al., 2014; Zheng et al., 2016), this study supported the view that descriptive analytics can identify the root causes of problems by investigating the status and the history of the mobile network traffic. Equally, predictive analytics can be used to seek future occurrences in the mobile network traffic by using the network event data (Atzmueller et al., 2016; Deka, 2014; Spieß et al., 2014). Likewise, prescriptive analytics can be used for optimization purposes to enhance network planning and allocation of network resources (Zheng et al., 2016).

Furthermore, the validity of the proposed framework can be tested by comparing the results obtained in the predictive analytics phase with previous studies (Diaz-Aviles et al., 2015). For instance, the study by Diaz-Aviles et al. (2015) used data feeds and logs of customer care calls gathered from a major African telecommunication company to predict user experience in real-time through a supervised learning approach and training of the restricted random forest model. The study supported the view that the dataset can be gathered by installing a probe in the MNO's network traffic (Diaz-Aviles et al., 2015). The datasets used by Diaz-Aviles et al. (2015) was low-level summary information using user-centric internet measurement for different aggregation time periods. Thus, it is possible to observe the most congested and less congested areas, which can lead to a larger number of Internet users' calls from areas that suffer high percentages of retransmissions (Diaz-Aviles et al., 2015). The data exploration observed by Diaz-Aviles et al. (2015) showed a promising correlation between the data feeds gathered from the network traffic and the registered calls to the care center, which enabled the prediction of user experience in real-time. Evidently, the restricted random forest showed 59% precision by Diaz-Aviles et al. (2015), representing a fair MOS score (Demirbilek & Gregoire, 2016). The low precision observed by Diaz-Aviles et al. (2015), indicated the unbalances observed in the data, because only a limited number of users would call customer care to report issues observed in the usage of the mobile Internet.

Overall, the proposed framework was envisaged to overcome the drawbacks observed in the study of Diaz-Aviles et al. (2015) by using expected variable values defined in the SLA. Equally, to avoid the unbalanced dataset observed in the study of Diaz-Aviles et al. (2015), historical customer care reports can be used to model user experience. This will enable the usage of historical customer care reports and historical user behavior to build personalized models for different segments of users and predict the perceived QoE more accurately. In view of all that has been mentioned so far, the MNOs can use the proposed framework for proactive purposes in the network traffic, to anticipate network problems and improve the overall mobile Internet customer experience in the telecoms industry.

METHODOLOGICAL INSTANCES OF THE PROPOSED FRAMEWORK

The proposed framework consists of three different phases: Data collection, Data preparation and Data modelling. In the data collection phase, this study assumed the mobile network traffic comprised of different types of datasets consisting of the three types of QoE influence factors. These datasets gathered from the mobile network traffic through active or passive probes injected into the network traffic. The gathered datasets can be referred to as big data if they constitute the big data characteristics (Volume, Velocity, Veracity, Value and, Varieties). For instance, in the case of the system data, the types of data expected to be collected are the values of download bit rates, upload bit rates, total bytes downloaded in the last 24 hours, hourly average number of retransmitted packets, maximum time needed by the user to receive the first byte from an application in the last 24 hours, minimum download time experienced by the user in the last 24 hours, minimum upload time experienced by the user in the last 24 hours, minimum hourly averaged round trip time in the last 24 hours, minimum hourly-averaged upload throughput, minimum hourly-averaged download throughput and many more based on the Internet application used by the users (Diaz-Aviles et al., 2015). Table 3 and Table 4 depict the example of the data attributes for HTTP and FTP respectively. An example of the context data can be in the form of the time of the day, date, longitude and latitude that can be used to indicate the context of the mobile internet users. While the examples of the user data come in the form of age, educational background and gender depending on the platform in which the data is collected. Moreover, some studies also argued that datasets comprising the subscription type and cost can also be collected from the network traffic to achieve the fairness criterion among the users (Xu, Xing, Perkis, & Jiang, 2011). Evidence have shown that some customers may have the same data rates, but a customer who has experienced a data rate increase may perceive greater experience (Rugelj et al., 2014). The fairness criterion will assist the MNOs to achieve an equilibrium level of an estimated perceived QoE as the sensitivity of customers tends towards infinity (Kim, Ko, & Kim, 2015). Therefore, the data collection is a very important phase when considering the modelling of the perceived QoE through the big data analytics approach. The types and quality of data collected from the mobile network have a huge influence on the result estimated or predicted by the perceived QoE. Once the data has been successfully gathered from the mobile network traffic, the next phase of the proposed framework was the data preparation phase.

Table 3

Http Datasets Attribute

Features gathered from Mobile Network
Time
Date
Latitude and longitude
Throughput total
Carrier 1 channel quality indicator (CQI) – Mean
Categorized received signal code power (RSCP): A1

Table 4

Ftp Datasets Attributes

Features gathered from Mobile Network
Time of the day
Date
Latitude and longitude
Throughput total (kbps)
Attachment set-up time (secs)
Application layer throughput downlink (kbps)
FTP download transfer time (seconds)
Connection duration (seconds)
FTP download throughput mean (kbps)

The second phase which is the data preparation phase involves data pre-processing, data exploratory analysis, feature selection and extraction from the big dataset collected from the mobile network traffic. The data pre-processing involves the cleaning, integration, and transformation of the data to suit the predictive analytics stage of the perceived QoE. Exploratory data analysis employs statistical techniques to aid and understand the dataset to be used for the predictive analytic stage. Feature selection aimed at selecting the most relevant attributes, while extraction combines the attributes into a reduced set of features. Hence, the feature selection and extraction enable the selection of subsets of features that are useful to build a good predictor, especially when some of the attributes are redundant. In most cases data preparation and descriptive analytics works together to enable a better understanding of the big dataset and prepare it for the modelling stage.

The third phase is the modelling phase and it consists of predictive analytics and prescriptive analytics. Predictive analytics involves the process of modelling perceived QoE or MOS. This phase comprises the observation of data instances. Observation of the data instances in this case represents the independent variable (that is, the extracted features from the big datasets and expectations) while the categories predicted are the possible values of dependent variables (perceived QoE) which are the classes or outcomes. The categorical outcome is usually represented as Excellent =5, Good =4, Fair =3, Poor =2, and Bad =1 (Demirbilek & Gregoire, 2016). The modelling of perceived QoE using the machine-learning algorithms would map the combination of input parameters to a class value to build an efficient model that

classifies extracted features with maximum precision through the perceived QoE function described as $QoE := f(User, Service, Variable)$ as used in the DQX model (Tsiaras & Stiller, 2014).

The DQX model was implemented experimentally by C.Tsiaras et al. (2015) using 34 volunteers (subjects), but a small number of subjects may not fully be ideal to generate a representative data. The DQX model assumed that every service (such as VOIP, Web-browsing, video streaming and skype voice calls) consists of both technical and non-technical QoE influence factors (system, context and human) that affect the perceived QoE. In this case, it is possible to distinguish two different types of variables which are, an increasing variable which increases as the users' experiences increases and a decreasing variable that decreases as the user' experience declines (Tsiaras & Stiller, 2014). This implies that, for every variable, there existed a certain value at which the user would experience a satisfactory perceived QoE with the service (Tsiaras & Stiller, 2014; C.Tsiaras et al., 2015). The stated values represent the expected variable values and are defined in the SLA (Tsiaras & Stiller, 2014; C.Tsiaras et al., 2015).

Overall, the idea of the DQX model can be extended to the predictive analytics phase in the proposed framework to determine the minimum and maximum perceived QoE based on the expected variable values defined in the users' expectation as stated in SLA. Expectation in relation to SLA is very important for modelling the perceived QoE. It allows the maximum, minimum and expected values to be defined for the QoE influence factors selected and extracted from the big datasets (Tsiaras et al., 2014). Equally, extending the view of the DQX model in the mobile environment scenario will enable the use of the expected values and influence factors of the individual variable to predict perceived QoE by considering multiple variables gathered from the mobile network traffic.

The prescriptive analytics takes advantage of the results obtained from both the descriptive and predictive analytics to decide the best decision or action that can be taken to improve the NP of the mobile network. As indicated in Figure 1, this article suggested the predicted MOS can be used for proper allocation network resources in locations where the MOS is below expectations. Conclusively, the use of big data analytics to manage the perceived QoE can enable the MNOs to provide proactive measures before the users would perceive any network distortion while using the Internet services provided by the MNOs. In addition, it can aid the MNOs to take optimal decisions for effective management of their NP to enable a better provision of the Internet services.

CONCLUSION AND FUTUREWORKS

This paper presented a proposed framework for modelling mobile Internet network perceived QoE through big data analytics approach. The proposed framework tended to overcome the specific context and service-related drawbacks associated with laboratory experiments. Likewise, the presented framework highlighted the importance of using datasets gathered from the mobile network traffic, as the datasets supported multiple context and service-related metrics for accurate modelling of the mobile Internet perceived QoE. In addition, methodological instances of the proposed framework were discussed, which can be used by the MNOs to effectively manage the network performance to aid a satisfactory QoE provision for mobile Internet users. Therefore, future work should implement the proposed framework by using the user experience datasets collected from the mobile network traffic. Equally, future work would validate the proposed framework to determine its applicability in real-life environments.

REFERENCES

- Alreshoodi, M., & Woods, J. (2013). Survey on QOE\QOS corelation models for multimedia services. *International Journal of Distributed and Parallel Systems (IJDPS)* 4(3), 53-72.
- Amour, L., Souihi, S., Hoceini, S., & Mellouk, A. (2015, August 23). A hierarchical classification model of QoE influence factors. *Wired/Wireless Internet Communications*, 225-238.
- Anchuen, P., Uthansakul, P., & Uthansakul, M. (2016). QOE model in cellular networks based on QOS measurements using neural network approach. *13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1-5). Chiang Mai: IEEE.
- Andrews, M., Cao, J., & McGowan, J. (2006). Measuring human satisfaction in data networks. *Proceedings of the 25th IEEE International conference on Computer Communications* (pp. 1-12). Barcelona, Spain: IEEE.
- Aroussi, S., & Mellouk, A. (2014). Survey on machine learning-based QoE-QoS correlation models. *IEEE*, 200-204.
- Aroussi, S., & Mellouk, A. (2016). Statistical evaluation for quality of experience prediction based on quality of service parameters. *23rd*

- International Conference on Telecommunications (ICT)* (pp. 1-5). Thessaloniki, : IEEE.
- Atzmueller, M., Schmidt, A., & Hollender, M. (2016). Data preparation for big data analytics: Methods and experiences. In M. Atzmueller, S. Oussena, & T. Roth-Berghofer (Eds.), *Enterprise Big Data Engineering, Analytics, and Management* (pp. 157-170). Hershey, PA: Business Science Reference. doi:10.4018/978-1-5225-0293-7.ch010
- Balachandran, A., Aggarwal, V., Halepovic, E., Pang, J., Seshan, S., Venkataraman, S., & Yan, H. (2014). Modeling web quality-of-experience on cellular networks. *MobiCom* (pp. 213-224). Maui, Hawaii, USA: ACM Doi: <http://dx.doi.org/10.1145/2639108.2639137>
- Barakovic, S., & Skorin-Kapov, L. (2013). Survey and challenges of QoE management issues in wireless networks. *Journal of Computer Networks and Communications*, 1-28.
- Barakovic, S., & Skorin-Kapov, L. (2015). Multidimensional modelling of quality of experience for mobile web. *Computers in Human Behavior*, 314 - 332.
- Barakovic, S., Barakovic, J., & Bajric, H. (2010). QoE Dimension and QoE measurement of NGN services. *18th Telecommunication Forum TELFOR*. Serbia Belgrade.
- Battisti, F., Carli, M., & Paudyal, P. (2014). QoS to QoE mapping model for wired/wireless video communication. *Euro Med Telco Conference (EMTC)* (pp. 1-6). IEEE.
- Calyam, P., Chandrasekaran, P., Trueb, G., Howes, N., Ramnath, R., Yu, D., Yang, D. (2012). Multi-resolution multimedia QoE models for IPTV applications. *International Journal of Digital Multimedia Broadcasting*, 1-3. Doi:10.1155/2012/904072.
- Charonyktakis, P., Plakia, M., Tsamardinos, I., & Papadopouli, M. (2016). On User-Centric modular QoE prediction for VoIP based on machine-learning algorithms. *IEEE Transactions on mobile Computing*, 15 (6) , 1443-1456.
- Chen, C. W., Chatzimisios, P., Dagiuklas, T., & Atzori, L. (2016). *Multimedia quality of experience (QoE) current status and future requirements*. United Kingdom : John Wiley & Sons, Ltd.

- Crawley, E., Nair, R., Rajagopalan, B., & Sandick, H. (1998). *A framework for QoS-based routing in the Internet*. USA: IETF RFC 2386.
- Deka, G. C. (2014). Big data predictive and prescriptive analytics . In P. Raj, & G. C. Deka, (Eds.), *Handbook of research on cloud infrastructure for big data analytics* (pp. 370-377). Herhely, USA: IGI Publishing.
- Demirbilek, E., & Gregoire, J. C. (2016, April 25). towards reduced reference parametric models for estimating audiovisual quality in multimedia services. *IEEE International Conference on Communications (ICC)*, (pp. 1-5). Kuala Lumpur: IEEE.
- DeMoor, K., Ketyko, I., Joseph, W., Deryckere, T., De Marez, L., Martens, L., & Verleye, G. (2010). Proposed framework for evaluating quality of experience in a mobile testbed- oriented living lab setting. *Mobile Networks and Applications*, 15 (3), 378-391.
- Diaz-Aviles, E., Pinelli, F., Lynch, K., Nabi, Z., Gkoufas, Y., Bouillet, E., Salzwedel, J. (2015). Towards real-time customer experience prediction for telecommunication operators. *IEEE International Conference on Big Data (Big Data)*, (pp. 1063-1072). IEEE.
- Du, H., Guo, C., Liu, Y., & Liu, Y. (2009). Research on relationship between QoE and QoS based on BP neural network. *Proceedings of IC-NIDC*, (pp. 312-315). IEEE.
- ESTI. (1994). *Network aspects (NA): General aspects of Quality of Service (QoS) and network performance (NP)*. France: ESTI Technical Report.
- ESTI TS 132410 V9.0.0, E. (2010). *Digital cellular telecommunication system (phase 2+); universal mobile telecommunication system (UMTS); LTE; telecommunication management; key performance indicators (KPI) for UMTS and GSM (3GPP TS 32.410 Version 9.0.0 release 9)*. European Telecommunication Union.
- ETSI EG 202 009-2 V1.3.1, E. (2014). *User group; Quality of telecom services; Part 2: User related indicators on a service specific basis*. European Telecommunications Standards Institute.
- Falk, T. H., & Chan, W.-Y. (2006). Single-ended speech quality measurement using machine learning methods. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6), 1935-1947.

- Fiedler, M., Hossfeld, T., & Tran-Gia, P. (2010). A generic quantitative relationship between Quality of Experience and Quality of Service. *IEEE Network, 24*(2), 36-41.
- Fiedler, M., Kilkki, K., & Reichl, P. (2009). From quality of service to quality of experience. *Executive Summary*. Dagstuhl, Germany.
- Geerts, D., De Moor, K., Ketykó, I., Jacobs, A., Van den Bergh, J., Joseph, W., De Marez, L. (2010). Linking an integrated framework with appropriate methods for measuring QoE. *QoMEX* (pp. 158-163). IEEE.
- Goleva, R., Atamin, D., Mirtchev, S., Dimitrova, D., & Grigorova, L. (2012). Traffic sources measurement and analysis in UMTS. *The Proceedings of the 1st ACM Workshop on High Performance Mobile Opportunistic System*, (pp. 29-32). Paphos, Cyprus.
- Gozdecki, J., Jajszczyk, A., & Stankiewicz, R. (2003). Quality of service terminology in IP networks. *IEEE Communication Magazine, 41*(3), 153-159.
- Hoßfeld, T., Biedermann, S., Schatz, R., Platzer, A., Egger, S., & Fiedler, M. (2011a). The memory effect and its implications on web QoE modeling. *Proceedings of the 23rd International Teletraffic Congress (ITC 2011)* (pp. 103-110). IEEE.
- Ibarrola, E., Xiao, J., Liberal, F., & Ferro, A. (2011). Internet QoS regulation in future networks: A user-centric approach. *ITU Kaleidoscope Conference* (pp. 148-155). IEEE Communication Magazine.
- ITU. (2014). *Measuring the information society report*. Geneva: International Telecommunication Union.
- ITU-T Recommendation. (2007). *Vocabulary for performance and Quality of Service- Amendment 1: New Appendix1-Definition of Qulaity of Experience (QoE)*. Geneva: ITU.
- ITU-T Recommendation E.800. (2008). *Definitions of terms related to quality of service*. ITU.
- ITU-T Recommendation P.10/G.100. (2016). *Amendment 5: New definitions for inclusion in recommendation ITU-T P.10/G.100; Vocabulary for performance and quality of service*. Geneva: ITU-T.

- Jumisko-Pyykko, Satu, & Vainio, T. (2010). Framing the context of use for mobile HCI: A review paper about mobile contexts use between 2000-2007. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 2(4), 1-28.
- Kim, H., Ko, S.-W., & Kim, S.-L. (2015). Network utility maximization with asymmetric user satisfaction. *IEEE Communication Letters*, 19(7), 1205-1208.
- Kotu, V., & Deshpande, B. (2015). *Predictive analytics and data mining concepts and practice with RapidMiner*. USA: Morgan Kaufmann .
- Laghari, K. u., Crespi, N., & Connelly, K. (2012). Toward total quality of experience: A QoE model in a communication ecosystem. *IEEE Communications Magazine*, 58-65.
- Laghari, K. u., Issa, O., Speranza, F., & Falk, T. H. (2012). Quality-of-Experience perception for video streaming services: Preliminary subjective and objective results . *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. Hollywood, CA, USA .
- Le Callet, P., Möller, S., & Perkiš, A. (2012, June 2). *Qualinet White paper on definitions of quality of experience output version of the Dagstuhl seminar 12181*. Retrieved from Qualinet.eu: http://www.qualinet.eu/images/stories/whitepaper_v1.1_dagstuhl_output_corrected.pdf
- Li, W., Spachos, P., Chignell, M., Leon-Garcia, A., Zucherman, L., & Jiang, J. (2016). Understanding the relationships between performance metrics and QoE for over-the-top video. *IEEE ICC, Communication QoS, Reliability and Modeling Symposium*. Doi:10.1109/ICC.2016.7511100: IEEE .
- Liotou, E., Tsolkas, D., Passas, N., & Merakos, L. (2015). Quality of Experience Management in mobile cellular networks: Key issues and design challenges. *IEEE Communications Magazine* , 145-153.
- Machado, V. A., Silva, C. N., Oliveira, R. S., Melo, A. M., Silva, M., Francês, C. R., Hirata, C. M. (2011). A new proposal to provide estimation of QoS and QoE over WiMAX Networks An approach based on computational intelligence and discrete-event simulation. *IEEE Latin-American Conference on Communications (LATINCOM)* . Doi: 10.1109/LatinCOM.2011.6107419 : IEEE.

- Menkovski, V., Cuadez-Sanchez, A., Oredope, A., & Liotta, A. (2009). Predicting Quality of Experience in multimedia streaming. *Proceedings of MoMM* (pp. 52-59). Kuala Lumpur, Malaysia: ACM.
- Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). *Big data imperatives. Enterprises big data warehouse, BI implementations and analytics*. New York: Apress.
- Möller, S., & Raake, A. (2014). Motivation and introduction. In *Quality of Experience: Advanced concepts, applications and methods* (pp. 3-10). Switzerland : Springer.
- Mushtaq, M. S., Augustin, B., & Mellouk, A. (2012). Empirical study based on machine-learning approach to assess the QoS/QoE correlation. *17th European Conference on Networks and Optical Communications (NOC)* (pp. 1-7). Vilanova i la Geltru: IEEE doi: 10.1109/NOC.2012.6249939
- Quintero, M. R., & Raake, A. (2011). Towards assigning value to multimedia QoE. *Third International Workshop on Quality of Multimedia Experience*.
- Raake, A., & Egger, S. (2014). Quality and Quality of Experience. In *Quality of Experience: Advanced concepts, applications and methods* (pp. 11-34). Switzerland: Springer.
- Raja, A., & Flanagan, C. (2008). Real-time, non-intrusive speech quality estimation: A signal-based model. *EuroGP*, 37–48.
- Reichl, P., Egger, S., Möller, S., Kilkki, K., Fiedler, M., Hossfeld, T., Asrese, A. (2015). Towards a comprehensive framework for QOE and user behavior modelling . *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, (pp. 1-6). Pylos-Nestoras : IEEE, Doi: 10.1109/QoMEX.2015.7148138
- Reichl, P., Egger, S., Schatz, R., & D’Alconzo, A. (2010). The logarithmic nature of QoE and the role of the Weber-Fechner Law in QoE assessment. *IEEE Communications Society*.
- Reichl, P., Tuffin, B., & Schatz, R. (2011). Logarithmic laws in service quality perception: Where microeconomics meets psychophysics and quality of experience. *Telecommunication System*, Springer. Doi: 10.1007/s11235-0119503-7

- Reiter, U., Brunnström, K., De Moor, K., Larabi, M.-C., Pereira, M., Pinheiro, A., Zgank, A. (2014). Factors influencing quality of experience. In *Quality of Experience: Advanced concepts, applications methods* (pp. 55-72). Switzerland : Springer .
- Rugelj, M., Volk, M., Sedlar, U., Sterle, J., & Kos, A. (2014). A novel user satisfaction prediction model for future network provisioning. *Telecommunication System*, 56 doi: 10.1007/s11235-013-9853-4 Springer, 417-425
- Sackl, A., Masuch, K., Egger, S., & Schatz, R. (2012). Wireless vs wireline shootout: How user expectations influence quality of experience. *QoMEX*.
- Sackl, A., Seufert, M., & Hoßfeld, T. (2013). Asking costs little? The impact of tasks in video QoE studies on user behavior and user ratings. *4th International Workshop on Perceptual Quality of Systems*. Vienna, Austria.
- Shaikh, J., Fiedler, M., & Collange, D. (2010). Quality of Experience from user and network perspectives. *Ann. Telecommun DOI: 10.1007/s12243-009-0142-x*, Springer, 47-57.
- Sharma, D., Meredith, L., Lainez, J., & Barreda, D. (2014). A non-intrusive PESQ measure. *Global SIP, perception inspired multimedia signal processing techniques* (pp. 975-978). IEEE.
- Singh, A., Mahmoud, A., Koensgen, A., Li, X., Göerg, C., Kus, M., Grigutsch, J. (2013). Enhancing Quality of Experience (QoE) assessment models for web traffic. *5th International Conference* (pp. 202-215). Cork, Ireland: Mobile Networks and Management, Springer.
- Skorin-Kapov, L., & Varela, M. (2012). A multi-dimensional view of QoE: the ARCU model. In *Proceedings of the 35th International Convention MIPRO* (pp. 662-666). Opatija, Croatia: IEEE.
- Spetebroot, T., Afra, S., Aguilera, N., Saucez, D., & Barakat, C. (2015). From network-level measurements to expected Quality of Experience: The Skype use case. *IEEE International Workshop on Measurement and Networking (M&N)*. Coimbra, Portugal: IEEE.
- Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., & Philippart, L. (2014). Using big data to improve customer experience and business performance. *Bell Labs Technical Journal* 18(4) doi: 10.1002/bltj.21642, 3-17.

- Stankiewicz, R., & Jajszczyk, A. (2011). A survey of QoE assurance in converged networks. *Computer Networks*, 1459-1473.
- Streijl, R. C., Winkler, S., & Hands, D. S. (2016). Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives. *Multimedia System*, 22 (2), 213-227.
- Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*. doi:10.1186/s40537-015-0030-3, 1-32
- Tsiaras, C., & Stiller, B. (2014). A deterministic QoE formalization of user satisfaction demand (DQX). *Annual IEEE Conference on Local Computer Networks* (pp. 227-235). Canada: IEEE.
- Tsiaras, C., Rösch, M., & Stiller, B. (2015). VoIP-based calibration of the DQX Model. *IFIP Networking Conference (IFIP Networking)*, Doi: 10.1109/IFIPNetworking.2015.7145309 . IEEE .
- Tsiaras, C., Sehgal, A., Seeber, S., Dönni, D., Stiller, B., Schönwälder, J., & Rodosek, G. D. (2014). Towards evaluating type of service related Quality-of-Experience on mobile networks. *7th IFIP on Wireless and Mobile Networking Conference (WMNC)*. IEEE.
- Tsolkas, D., Liotou, E., Passas, N., & Merakos, L. (2016). A survey on parametric QoE estimation for popular services. *Journal of Network and Computer Applications*.
- Vuckovic, P. M., & Stefanovic, N. S. (2006). Quality of Experience of mobile services . *14th Telecommunications Forum TELFOR* . Serbia, Belgrade.
- Wechsung, I., Schulz, M., Engelbrecht, K.-P., Niemann, J., & Möller, S. (2011). All users are (not) equal- The influence of user characteristics on perceived quality, modality choice and performance. *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*.
- Xu, J., Xing, L., Perkis, A., & Jiang, Y. (2011). On the properties of mean opinion scores for Quality of Experience management. *International*

Symposium on Multimedia (pp. 500-505). Doi: 10.1109/ISM.2011.88:
IEEE Computer Society .

Yang, K., Liu, R., Sun, Y., Yang, J., & Chen, X. (2016). Deep network analyzer (DNA): A big data analytics platform for cellular networks. *IEEE Internet of Things Journal*, 1-9.

Zheng, K., Yang, Z., Zhang, K., Chatzimisios, P., Yang, K., & Xiang, W. (2016). Big data-driven optimization for mobile networks towards 5G. *IEEE Network*, 44-51.