Quality of Experience of VoIP Service: A Survey of Assessment Approaches and Open Issues

Sofiene Jelassi, Gerardo Rubino, Hugh Melvin, Habib Youssef, and Guy Pujolle

***Abstract*—This survey gives a comprehensive review of recent advances related to the topic of VoIP QoE (Quality of user' Experience). It starts by providing some insight into the QoE arena and outlines the principal building blocks of a VoIP application. The sources of impairments over data IP networks are identified and distinguished from signal-oriented sources of quality degradation observed over telecom networks. An overview of existing subjective and objective methodologies for the assessment of the QoE of voice conversations is then presented outlining how subjective and objective speech quality methodologies have evolved to consider time-varying QoS transport networks. A description of practical procedures for measuring VoIP QoE and illustrative results is then given. Utilization methodology of several speech quality assessment frameworks is summarized. A survey of emerging single-ended parametric-model speech quality assessment algorithms dedicated to VoIP service is then given. In particular, after presenting a primitive single-ended parametric-model algorithm especially conceived for the evaluation of VoIP conversations, new artificial assessors of VoIP service are detailed. In particular, we describe speech quality assessment algorithms that consider, among others, packet loss burstiness, unequal importance of speech wave, and transient loss of connectivity. The following section concentrates on the integration of VoIP service over mobile data networks. The impact of quality-affecting phenomena, such as handovers and CODEC changeover are enumerated and some primary subjective results are summarized. The survey concludes with a review of open issues relating to automatic assessment of VoIP.**

***Index Terms*—Quality of users' Experience (QoE), Speech quality assessment (SQA), VoIP, Parametric SQA algorithms, Mobile heterogeneous networks.**

I. Introduction

THERE is an increasing recognition that the quality of network services (QoS) should be evaluated according to their Quality of Experience (QoE) rather than the classical network-oriented metrics, such as delay, availability, response time, jitter, echo, and packet loss [1], [2]. Indeed, the interpretation as good or bad of a QoS metric can be confusing because it ignores its concrete effect on clients [3], [4]. Moreover, the QoE can be deemed as fine in spite of a slow response time, high jitter, or packet loss [3]. The significance of the QoE arena resides in its prospective considerable economic and engineering consequence on the worldwide transport infrastructure design and services' deployment and consequent revenue [5]. This explains the extreme attention given by the ITU standardization bodies to the QoE issue and its related branches. In particular, the ITU-T Study Group 12 (SG12) is focused on Performance, QoS and QoE, and defines a large number of high priority questions that are being addressed in collaboration with several ICT (Information and Communication Technologies) intra- and inter- related partners, e.g., SG16, SG13, ETSI, 3GPP, and IETF [6]. Such questions require detailed investigation using innovative ideas and research.

The ITU defines QoE as: "A measure of the overall acceptability of an application or service, as perceived subjectively by end-user". Precisely, the QoE measure considers both the services' and users' quality influencing factors [5]. The service factors include availability, reliability, set-up and response times, type of terminals, etc. The users' factors include emotions, experience, motivation, and goals. The QoE measure has a distinct meaning according to the specificity of each application. For example, a positive QoE measure in the context of a voice conversation signifies that the call is characterized by an excellent voice *transmission quality* and *ease of communication*. However, a positive QoE measure in the context of a Web surfer signifies that good quality graphics and pictures are downloaded within an acceptable timeframe. Similar observations can be drawn for other network services, such as IPTV, Video conference, multimedia messaging, MMOG, VoD, and E-commerce.

In this survey, we narrow the scope to VoIP QoE , a service that has recently gained significant market penetration [7]. The technical reports of TeleGeography indicate that PC-to-PC and PC-to-PSTN calls are now very significant in context of overall international voice market [8]. In reality, PC-to-PC and PC-to-PSTN services are replacing and extending conventional telephone systems in homes and enterprises, a move referred to as *network service convergence* [9]. The basic building blocks installed at sender and receiver sides that are required for running a PC-to-PC VoIP service are presented in Figure 1. As we can see, the processing pipeline of speech wave signals starts at the speech encoder that is responsible for the acquisition, sampling, and compression operations. Precisely, the captured analog voice signals are sampled with a fixed frequency, where each sample is encoded. The encoding mechanism aims to balance conflicting require-

Manuscript received 27 May 2010; revised 7 December 2010. The work described in this paper is partially supported by the national French project ANR VERSO ViPeer.

S. Jelassi and G. Rubino are with Institut National de Recherche en Informatique et Automatique (INRIA), centre de Rennes, Bretagne Atlantique, France (e-mail: sofiene.jelassi@gmail.com, gerardo.rubino@inria.fr).

H. Melvin is with Discipline of Information Technology, National University of Ireland (NUI), Galway, Ireland (e-mail: Hugh.Melvin@nuigalway.ie).

H. Youssef is with Higher Institute of Computer Science and Communication Technologies (ISITC), Sousse, Tunisia (e-mail: Habib.Youssef@fsm.rnu.tn).

G. Pujolle is with Laboratoire d'Informatique de Paris 6 (LIP6), Université de Pierre et Marie Curie (UPMC), Paris, France (e-mail: Guy.Pujolle@lip6.fr).

Digital Object Identifier 10.1109/SURV.2011.120811.00063

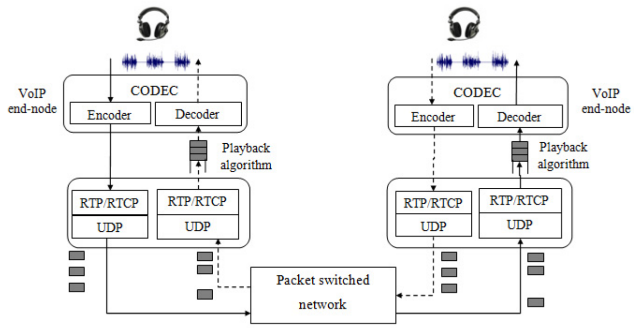


Fig. 1. Pipeline processing of interactive PC-to-PC voice conversations.

ments of reduced bandwidth while keeping a good perceived quality. The encoding scheme suitable for interactive voice conversations can follow either signal- or parametric model-based coding methodologies [10]. The signal-based encoding strategy directly operates on wave signals in the time-domain. The idea behind the signal-based compression schemes stems from the observation that the difference between consecutive voice raw samples is typically small. This means that the difference necessitates a lower number of encoding bits than those required for encoding the sample itself. On the other hand, parametric model-based strategy extracts relevant signal characterization measures of a given block of voice samples. The encoded characterization' parameters are sent to the receiver side that can then artificially *synthesize* the original speech wave.

The *bitstream* produced by a given speech encoder is placed into packets following a specific filling mechanism in order to ensure the compatibility between heterogeneous devices. Each speech encoder defines the packetization mechanism that aims at achieving a trade-off between end-to-end delay and packet rate. At the receiver side, a playback algorithm determines the playback time of each received speech frames. In addition to voice data transport channels, there is a signaling channel that is responsible for, amongst others, caller and callee identification, call redirection, presence service, edge-device configuration, and QoS negotiation [11].

VoIP QoE is sensitive to network delay variation, referred to as *network delay jitter*. This stems from the fact that received media packets are presented at run-time [7]. As such, a media packet that arrives after its expected deadlines is discarded. Erasing some VoIP packets is tolerable to a certain extent, unlike legacy reliable data network services, such as remote control, file transfer, and e-mail. These reasons explain why voice packet streams are often carried using the unreliable transport protocol UDP. In such a case, the periodically injected voice packets may undergo transmission impairments such as excessive one-way network delay, jitter, packet loss and reordering. These are the main sources of impairments that significantly degrade the QoE of VoIP service. There are multiple remedies that have been developed to overcome the negative perceived effects of such sources of quality degradation. The existing mechanisms can be classified as follows:

- *Network-centric strategies*: These enhance the QoE through the integration of suitable QoS mechanisms within the network to satisfy the specific needs of delay-sensitive services. In such a case, intermediate nodes can perform call admission, preferential treatment of received packets, or policy enforcement [12]. This requires an upgrade or replacement of all existing nodes, which is a significant challenge, particularly when VoIP calls are routed across heterogeneous, large scale, and geographically distributed transport systems.
- *Application-centric strategies*: These seek to improve the QoE through the deployment of advanced control mechanisms at sender and receiver sides to smartly deal with sources of quality degradation observed over data networks. For example, the sender can adapt its transmission rate, packet loss protection strategy, packetizing approach, and packet transfer scheduling according to the prevailing features of delivery pathway. On the other hand, the receiver can absorb network delay jitter through optimal jitter buffer strategies and apply certain mechanisms for packet loss concealment.

The prevailing research trend consists of integrating network and application QoS enhancing strategies for the improvement of the QoE of VoIP under arbitrary network conditions. This can be seen by the tremendous attention given by the research community to cross-layering architectures [13]. In such a case, all forwarding nodes will be aware of the specific features and needs of each crossing flow. Suitable actions will be taken by each node in distributed and autonomic ways that enable efficient management of a large variety of services delivered over a heterogeneous, mobile, and time-varying QoS and topology transport system. On the other hand, the applications will be highly resilient to cope with time-varying QoS transport network situations characterized through a sophisticated learning process [14], [15]. Hence, they will be able to smartly adapt their behavior following the prevailing network in order to optimize the QoE.

This paper addresses the topic of perceptual quality assessment of voice calls over packet-based time-varying QoS transport networks. It surveys both subjective and objective speech quality assessment methodologies, together with their potential applications, such as system tuning and diagnosis. Moreover, it describes the deployment and exploitation methodologies of objective SQA algorithms. In particular, it details single-ended parametric SQA[1] algorithms characterized by their ability to assess at run-time the perceived quality using a set of measures gathered from the header content of received packets. These pertinent features suit several applications, such as management, diagnosis, and billing. We review the emerging single-ended SQA algorithms that subsume the effect of time-varying QoS offered by data networks, transient loss of connectivity, and the unequal perceived importance of presented speech waves. Further, we identify and discuss a set of open issues yet to be addressed, for accurate quality estimation of QoE of VoIP calls over heterogeneous networks. These include the perceived effect of dynamic jitter buffer policies, which result in the alteration of the temporal structure of the original voice

[1]The term SQA (Speech Quality Assessment) will be used hereafter to refer to a process that leads to quantify the QoE metric in the context of VoIP service from a given perspective.

stream, and the need to elucidate and to better quantify the QoE over mobile networks.

The remainder of this paper is organized as follows. Section 2 summarizes subjective and objective speech quality assessment (SQA) methodologies reported in the literature. Section 3 describes utilization methodologies of SQA algorithms. Section 4 presents proposed single-ended SQA algorithms to assess at run-time the QoE of voice calls over data networks. Section 5 reviews SQA methodologies over mobile networks. Section 6 discusses some new issues related to the assessment of QoE of adaptive VoIP applications. Section 7 concludes the paper.

II. A GUIDE TO SPEECH QUALITY ASSESSMENT METHODOLOGIES

The evaluation of QoE of voice conversations can be performed using *subjective trials or objective SQA algorithms* [16]. In the next sections, we will substantially discuss subjective and objective quality assessment methodologies.

A. Subjective-driven speech quality

Subjective methodology quantifies users' satisfaction (or perceived quality) using a panel of human subjects. In such a case, clean speech material should be prepared following reference conditions, namely ambient noise, quantization level, sampling rate, sample precision, stimulus duration, structure, and content, microphone and speaker acoustic features, etc. The reference speech sequences are injected in an existing, a simulated, or an emulated transport system that generates at the opposite side degraded / processed speech sequences. Each condition is evaluated using several speech sequences spoken by multiple male and female speakers. Form a practical viewpoint, existing subjective SQA methodologies can be classified as Laboratory- and Internet- based approaches that will be presented in the following sections.

1) Laboratory-based subjective SQA approaches: Undoubtedly, the laboratory-based subjective SQA strategies defined in the ITU-T Rec. P.800 are the most popular among practitioners [17]. In such a case, the selection of a subject panel is a critical aspect for SQA accuracy and consistency. The ITU-T Rec. P.800 indicates that more than six human subjects should participate during a subjective testing session. Human subjects are randomly selected from the ordinary population that habitually use telephone communication. Moreover, they should not have been directly involved in a subjective essay during the last six months. The speech materiel should be unknown to the participants. Note that an analytical description of laboratory-based subjective SQA trials is detailed in [18].

The ITU-T Rec. P800 specifies several subjective approaches (see Figure 2) to quantify the quality of degraded speech sequences according to various perspectives and goals [17]. Basically, we can classify defined subjective strategies into two main categories:

- *Absolute-based rating methods*: In this approach, subjects are asked to rate speech quality or detectibility based solely on the distorted speech sequences. The judgment is performed using a standard scale that is specific to each assessment method. In practice, conducted subjective tests mostly follow the ACR (Absolute Category Rating) method that quantifies the *perceived speech quality*, rather than the annoying degree. In such a case, subjects manipulate a 5-point scale where 1 (resp. 5) refers to a bad (resp. toll) quality. At the end of each test, a Mean Opinion Score (MOS) is calculated that quantifies the perceived speech quality under such a condition. To quantify the detectibility of sources of distortion, such as echo, reverberation, side-tone, Quantal-Response Detectability (QRD) tests are more suitable and should be followed [17].
- *Comparison-based rating methods*: In this approach, subjects listen to both reference and degraded speech sequences. Comparison-based rating methods have been proposed to enable subtle assessment of voice channels and CODECs that produce slightly distinct good perceived qualities. The ITU-T Rec. P.800 describes two comparison-based rating methods referred to as DCR (Degradation Category Rating) and CCR (Comparison Category Rating). They can be distinguished according to the temporal presentation policy of manipulated speech sequences and their assessment scale. Notice that DCR and CCR methods are mainly intended for the evaluation of the perceived effects of various kinds of noises and the efficiency of noise removal devices. With DCR, four pairs of reference speech signals immediately followed by the degraded one are presented to a set of listeners. Subjects are then asked to assess the degree of degradation of distorted speech sequences with respect to reference ones. The produced scores of DCR trials are referred to as DMOS. With CCR, four pairs of reference and distorted speech sequences are presented *in an arbitrary order* unknown by the listeners. Subjects are then asked to evaluate the degree of improvement or degradation of speech quality of presented speech sequences. The produced scores of CCR are referred to as CMOS. Notice that unlike DCR, CCR enables quantification of speech treatment that improves speech quality of the reference signal.

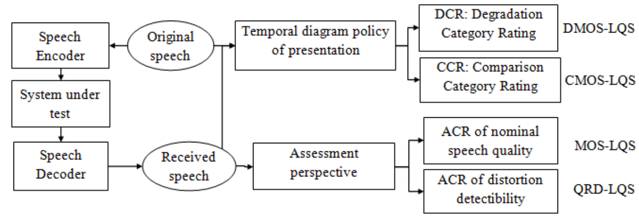


Fig. 2. Taxonomy of ITU-T Rec. P.800 listening subjective approaches.

At their inception, subjective listening tests have been designed to assess speech quality delivered over transport system that only involves time-invariant impairment sources. This explains why the duration of presented stimulus has been confined to 8-20 sec. that easily reflects the perceived quality throughout the entire call. This assumption is reasonable for legacy telecom networks, but is a poor match for emerging advanced mobile and IP transport systems that offer a time-varying QoS during a voice session. This limitation has led to new subjective methods to quantify speech quality delivered through time-varying QoS channels [19]. In such a situation, it is needful to distinguish between *instantaneous* and *overall* perceived qualities. The instantaneous perceived quality refers to the user's satisfaction at a specific instant during the presentation of a time-varying speech quality sequence. The overall perceived quality refers to the user's satisfaction at the end of a subjective essay. As such, human short-term memory comes into play. Notice that ITU-T Rec. P.880 gives a comprehensive explanation on how to perform Continuous Evaluation of Time-Varying Speech Quality, abbreviated as CETVSQ [19]. Precisely, the instantaneous perceived quality is assigned by subjects using a slider attached to an absolute graphical MOS scale ruler. A new score of instantaneous quality is recorded every 500 ms. At the end of a test, subjects give an overall score that quantifies their degree of satisfaction about the whole experience. In such a continuous evaluation essay, the durations of stimulus used to quantify time-varying speech quality vary between 45 sec and 3 min, in opposition to 8-20 sec stimulus duration used for legacy subjective trails. Notice that the continuous evaluation of time-varying speech quality have revealed the existence of a recency effect which signifies that subject rating is linked to the degradation temporal location with respect to the overall sequence [20].

Subjective trials can also be used to quantify the *conversational quality*. The listening quality trials are used to quantify the *hearing* quality of distorted speech sequences. The conversational quality trails are used to quantify the quality of interactivity experienced by end users. Such experiences include the effect of delay-related impairments, such as one-way delay and echoes. Basically, the higher the one-way delay is, the poorer is the quality of interactivity and the greater is the negative perceived effect of echoes[2] [21]. This results in a significant reduction of quality of interaction even in presence of good echo cancellation. During such an experience, subjects are placed in a cabinet and asked to follow a given conversation scenario subject to an introduced degree of impairments [16]. At the end of a conversation, subjects are asked to vote the overall perceived quality. Notice that different conversational tasks can greatly impact on perceived quality. This is related to the kind of conversations that can be urgent, professional, familial, or informal [22].

Recently, ITU-T P.800 subjective test methodology has been criticized by several researchers [4], [23], [24]. In fact, subjective tests following ITU-T Rec. P.800 should be conducted in an isolated sound room, using a dedicated equipment, and suitably selected subject panel. This is an important challenge in academic and industrial contexts and explains why subjective trials are confined to standardization institutes and a few powerful telecom operators. The requirement for an isolated sound room is dictated by the need to assess, in a controlled way, speech CODECs, echo cancellers, and noise reduction algorithms that produce slightly different perceived qualities, which are difficult to differentiate under realistic operational circumstances. In addition, isolated sound room enables unbiased rating of the effect of signal-related impairments, such as quantization, the physical characteristics of the room, and circuit noises, acoustic features of phones, and echoes. This is generally important in telecommunications, but, becomes less significant for the assessment of packet-based VoIP conversations [25]. Indeed, in this latter case, new sources of impairments are more critical to perceived quality, such as packet loss, CODEC switching, packetizing strategy, temporal impairments which affect the temporal structure of the original sequence, and intermittent loss of connectivity. It is worth to note here that the principal advantage of ITU-T Rec. P.800 resides in the fact that it ensures uniformity between subjective scores coming from different laboratories. This explains why it is nowadays used for the evaluation of the perceived quality of speech transmission quality over VoIP, albeit, ITU-T Rec. P.800 has been intended for the assessment of voice conversations delivered over a telecom transport infrastructure.

2) Internet-based subjective SQA approaches: To properly and efficiently assess VoIP transmission quality, Internet-based subjective approaches have been developed in the literature [23], [24]. Note that the specification of legacy subjective test methodology was done when the prevailing multi-service Internet was in its early age. Moreover, during the early days of the Internet, VoIP-similar services were very much a secondary objective. The advantage of using the Internet or a local area network (LAN) can be summarized as follows:

1) It allows for subjects to perform the quality assessment in realistic conditions. Indeed, in such a case, subjects do their judgement in familiar environments, such as an office or a quite room. As such, subjective tests are more comfortable and they can be done in a personalized fashion according to the pace and preferences of each subject.
2) It enables a subjective test campaign that involves subjects located in geographically distributed locations. The subjects can perform their judgement according to their preferred time slot. To accelerate the procedure, it can be useful to gather subjects in a given laboratory connected to the Internet, where every subject uses a multimedia computer and head set to perform his judgement.
3) It enables for the administrator a simple and efficient management of subjects, speech material, and recorded data.

Basically, we can distinguish two ways to conduct subjective experiences using the Internet that we call the streaming and download modes:

- *Subjective quality assessment following streaming mode*: A set of high-quality speech sequences are placed on a streaming server. Subjects are asked to playback speech sequences at real-time using an arbitrary player installed on their computers. After hearing the whole speech sequence, subjects give a MOS score that quantifies their satisfaction. Finally, the assigned score, a set of characterizing parameters, such as ambient S/N ratio, one-way delay, jitter, and information about packet loss process, and configuration parameters are sent back toward the streaming server. Notice that in such a case,

[2] Beyond a certain one-way delay threshold, an echo removal device should be set-up.

speech sequences are distorted in an arbitrary way, thus facilitating assessment of the perceived quality of live networks. Notice that streaming strategy requires a permanent connection to the Internet. Subjects should immediately listen to played speech sequences and give their assessment.

- *Subjective quality assessment following download mode*: A set of high-quality and degraded speech sequences are placed on a Web server. Subjects are instructed to download the whole sequence using a reliable transport protocol. Next, they are asked to play recorded sequences on their computer in a given order. Finally, they are asked to give a score about perceived quality. This information is sent back toward the server. Basically, most of Internet-based quality assessment experience follows this strategy since it is simpler to implement than streaming strategy. Moreover, the quality assessor has the ability to control the intensity of impairment sources.

In practice, the selection of an assessment strategy depends on the intended goals and objectives. For example, streaming mode can be utilized for the diagnosis and tuning of an existing transport system. However, the download mode can be utilized for the development of parametric quality assessment models for new speech CODECs, packet loss concealment schemes, de-jittering policies, noise removal devices, etc. In summary, Internet-based techniques enable more efficiency, realism, wide access and ease of management. It reduces significantly the expensive cost and time of ITU-T P.800 subjective test methodology. Nevertheless, the main drawback of Internet-based subjective test is the lack of a controlled operation environment where subjective trials are conducted, e.g., very low background noise, headset quality, personalized settings of media players, etc. They thus provide some kind of average assessment, the average taken over a large set of possible environments. To help reduce biased rating, subjects are asked to provide additional information about their operational situations to account for it during result analysis.

Besides the cumbersomeness and expense of legacy laboratory-based subjective trails, the absolute scale used to judge perceived quality by subjects constitutes a principal source of confusion [4], [24]. For example, subjects can interpret the label "Poor" and "Good" quality in different ways, despite having a similar level of experience in a given assessment test. Moreover, since subjects are asked to only assign a single score at the end of an experience, it will be difficult to rigorously distinguish the perceptual importance of each presented speech segment. This explains why a large number of subjects are needed to gather sufficient information before drawing consistent statements. Furthermore, reduced concentration and tiredness of subjects after multiple repetitions of subjective trials complicate the process.

To overcome such shortcomings, a new Internet-based framework for measuring network Quality of Experience (QoE), known as OneClick, has been developed [24]. The term OneClick has been chosen since users are only asked to click a dedicated button whenever they feel dissatisfied with the quality of the application in use. As such, OneClick is *intuitive* because subjects are not asked to discriminate between "Good" and "Fair" or between "Bad" and "Poor". Moreover, in contrast to legacy subjective experience, OneClick enables sending an immediate feedback which eliminates for the subject the need to keep in mind previous experience to perform their judgement at the end of a test. As Internet-based subjective trials, OneClick is *lightweight* because it does not need important logistics to perform subjective tests. In addition, OneClick is *efficient* since it captures a large number of click events from a subject in single test. As such, a single test of a few minutes is sufficient to identify the perceived quality of a subject over a wide range of network conditions. Furthermore, OneClick is *time-aware* since it enables revealing of behavior rating of subjects over time-varying QoS transport networks. The click rate is utilized as a judgement metric to compare perceived quality of distinguished network settings. Obviously, the higher the click rate is, the lower is the perceived quality. It is worth to note here that results obtained using Internet- and Laboratory -based SQA approaches are incompatible. The development of a bridging methodology that links both subjective assessment strategies is a prospective research arena.

Finally, it is critical that gathered subjective scores should be statistically treated to derive consistent results. ITU-T Rec. P.800 states that an ANOVA statistical analysis should be conducted to study the effects of introduced impairments on quality degradation and improvement [17]. Moreover, ITU-T Rec. P.800 indicates that the significance test should be performed. Graphical representations, such as a pie chart or a scatter plot may be useful for further explanation and study of produced subjective results.

B. Objective-driven speech quality

The subjective methodology is unable to rate at run-time the perceived quality of live voice calls which confines its utility to a limited range of applications, such as system tuning and diagnosis. Generally speaking, it is well-recognized that subjective trials are time consuming, cumbersome, and expensive [16]. This explains why objective methodologies which aim to quantify the perceived quality a using machine-executable SQA algorithm are an elegant and effective solution. Such objective SQA algorithms should be informed by a comprehensive analysis of subjective scores obtained using standardized subjective tests. Once an objective SQA algorithm is developed and calibrated, it can be faithfully utilized to automatically estimate the QoE under similar impairment conditions. Note that it is mandatory to resort to subjective approach to rationally quantify the effect of new uncounted for impairment phenomena and situations.

In the last decade, multiple SQA algorithms to estimate either absolute MOS score or conversational quality have been proposed and developed. The diagram given in Figure 3 gives taxonomy of existing SQA methodologies. They have been classified as function of their inputs as black-box signal-layer and glass-box parametric-model SQA approaches. The existing SQA algorithm that follows one of the two strategies are segregated following their operational mode [16]. The following sections will explain how speech quality is estimated using each methodology.

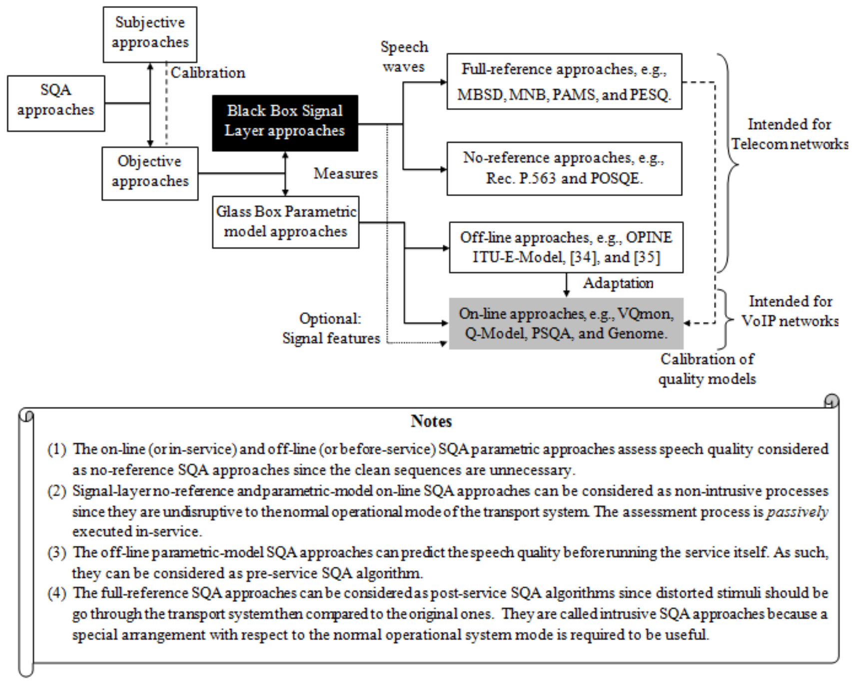


Fig. 3. Taxonomy of existing SQA methodologies.

1) Signal-layer black-box approaches: The signal-layer black box category of SQA algorithms estimates the perceived quality by processing speech waves without specific knowledge of the underlying transport network and terminals features. They can be classified as *full-reference* (also termed comparison-based, double-ended, intrusive) SQA algorithms because they need both the reference and degraded speech sequences, and *no-reference* (also termed single-ended, non-intrusive) SQA algorithms if they only process the degraded sequences to estimate the perceptual quality.

Basically, signal-layer comparison-based SQA algorithms divide original speech sequences into 50% overlapping frames of duration 20 ms. The reference and corresponding degraded frames are transformed into a psychoacoustic domain that considers the features of the human auditory system. Next, the audible distortion between reference and degraded speech frames is calculated and recorded. The gathered distortion values are combined to deduce the final score. The ITU-T defined two SQA algorithms following such a strategy in ITU-T Rec. P.861 and P.862 denoted respectively as MNB and PESQ [16], [26]. These algorithms are characterized by their good accuracy to estimate speech quality. However, their high time and space complexities are pertinent drawbacks. To decrease computing-complexity, there is a possibility for manipulating *elementary objectives voice quality metrics* that are able to capture distinct types of distortions [27]. S. Rein et al. review existing *elementary objectives metrics* for the sake of evaluation of voice quality over wireless data networks under the ROHC (Robust Header Compression) strategy [28]. They indicate that the parametric cepstral distance, which is the difference in shape of original and distorted spectra, correlate with objective measures. As such, they propose a 2-order polynomial function that maps measured cepstral distance to MOS domain [28]. Note here that it is expected that elementary objective metric methodologies result in a lower accuracy than psychoacoustic objective metrics.

Notice that comparison-based SQA algorithms synchronize reference and degraded speech sequences before processing since the comparison should be made between the original and the corresponding degraded speech frame. The synchronisation should be made on a frame-by-frame basis using treated speech signals themselves, i.e. without further control information about the temporal structure of processed speech sequences. Notice that such an alignment removes any temporal distortion introduced by adaptive jitter buffer policies. This constitutes a pertinent deficiency of prevailing comparison-based SQA algorithms that will be described in detail in Section 6.

No-reference signal-layer SQA algorithms extract directly distortion measures from degraded speech signals, e.g., additional noises, interruption, level saturation, pauses, and naturalness of speech sound that are adequately mixed to derive the final score [29]. Needless to say, the combination rule and the weighting factor of each metric are obtained through a training process applied on subjective MOS scores.

2) Parametric-model glass-box approaches: These quantify the QoE of voice calls through the full characterisation of the underlying transport network and edge-devices using a set of dedicated parameters, such as room, circuit and quantization noises, packet loss, coding scheme, delay jitter, one-way delay, jitter, and talker and listener echoes. Figure 4 gives an example of parameters that are required for the full characterization of an advanced Mobile and IP transport system. Suitable parameters, models, and combination rules should be selected and specified through extensive subjective testing. Note that parametric-model SQA algorithms are able to estimate users' satisfaction either on-line or off-line (, i.e., without considering live measures). Emerging on-line parametric-model SQA algorithms will be addressed and explained later in Section 4.

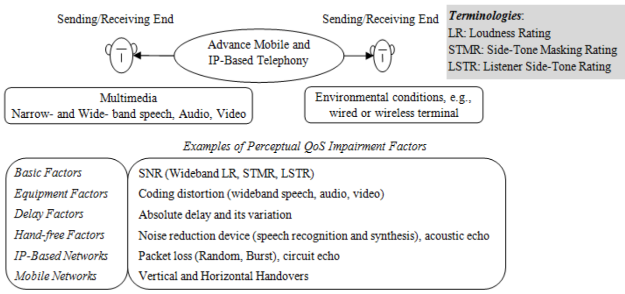


Fig. 4. Perceptual QoS factors affecting the perceived quality of advanced mobile and IP-based communication [31].

A well-known off-line parametric-model computational SQA algorithm, named E-Model, has been defined by the ITU-T in Rec. G.107. The E-Model predicts users' satisfaction under a given planned transport configuration [30]. The prevailing version of ITU-T E-Model has 21 input parameters that tightly characterize the designed configurations. The output of the ITU-T E-Model is a single scalar value referred to as *rating factor* R that lies between 0 (Worst Quality) and 100 (Excellent Quality). In practice, a designed transport configuration which results in a rating factor below than 60 should be avoided. In such a case, suitable remedies should be undertaken to improve users' satisfaction. The individual sources of quality degradation are classified into three primary impairment factors, which will be detailed in Section 4.1. ITU-T Rec. G.107 provides the range values of each individual parameter and the combination models that enable computing the value of each primary impairment factor. The final rating score is calculated through the key assumption that the perceived effects of primary impairment factors are additive on a psychological scale.

The E-Model has many limitations, especially in the context of many VoIP transport systems that need to be understood to avoid misuse. Within the E-Model, input parameters are constant during a voice session, which is inconsistent with the features of time-varying QoS transport networks. Moreover, it has been subjectively shown that applying an additive effect of primary impairment factors on psychological scale entails a significant inaccuracy of users' satisfaction. Moreover, ITU-T E-Model is speech-independent. This explains why the utilisation of ITU-T E-Model should be confined to planning purposes as defined by the ITU-T. In [32], the author developed more sophisticated combination rules of primary impairment factors to more accurately reflect users satisfaction. Moreover, a combination of E-Model and PESQ[3] (Perceptual Evaluation of Speech Quality) has been proposed to subsume the features of delivered speech waves [33]. The ITU-T E-Model is currently under significant revisions to account for the evolving features of IP-based NGN, such as wideband speech coding schemes, adaptive behavior of VoIP application, and specific features of each language [34], [35]. Notice that recently there is the emerging of the concept of *partial-reference* parametric SQA algorithms that quantify the QoE using network measurable parameters and metadata characterizing the original structure of reference speech sequence, such as the perceptual importance of sent voice frames and operational mode of voice activity detector [36], [37].

It is worth to note here that the selection of suitable SQA algorithms depends on the intended goals and the measurement context [31]. Intrusive double-ended signal-layer SQA algorithms are suitable for onsite diagnosis and tuning of terminals and networks. Off-line parametric-model SQA algorithms are suitable for planning purposes, i.e., during the design of communication systems. On-line parametric-models and signal-layer single-ended SQA algorithms are suitable for quality monitoring, billing, and service management. Typically, telecom operators with a R&D (Research & Development) division, such as Ericsson, Deutsch Telekom and British Telecom, develop their *propriety* implementations of previously described SQA algorithms and their companion software tools for measures' acquisition, recording, and analysis that satisfy their specific needs. However, the majority of telecom operators and voice service providers delegate the task of assessment of their transport infrastructure, services, and equipments to specialized corporations, such as GL, OPTICOM, Telchemy, and HEAD Acoustics. They are responsible for the evaluation, the detection of anomaly, and the proposal of efficient remedies for improving the perceived quality over the tested transport systems and services.

In the context of VoIP, online parametric-model single-ended SQA approaches (Grey Box in Figure 3) are more suitable because of their *reduced overhead* and their ability to manage and monitor live VoIP sessions delivered through heterogeneous transport infrastructures. In such a case, the measure of speech quality is performed through the monitoring of the flow of voice packets without any access to speech signals which is preferable for privacy reasons. However, generally speaking, parametric-model SQA algorithms are relatively less accurate than signal-layer SQA algorithms. Notice that signal-layer SQA algorithms have been extensively studied in the literature since early days of telecom networks [16]. This explains why this paper solely details new emerging *single-ended parametric-model SQA algorithms* of VoIP service, which are well-suited to accommodate stakeholder needs.

III. Utilization Methodologies and Outputs of Objective SQA Algorithms

In this section, we enumerate different deployment and exploitation methodologies of objective SQA algorithms. In their early days, objective SQA approaches were implemented using C-like languages, which needed expert people to manipulate them. During the last decade, new speech quality assessment frameworks have evolved, which are equipped with user-friendly interfaces. For instance, GL[4] Corporation has developed *proprietary* software called VQuad, which is an easy-to-use assessment platform of speech signals delivered over a wide range of heterogeneous networks, namely, TDM, VoIP, and wireless networks. VQuad enables automatic testing / simulating of PSTN and VoIP phones. Moreover, VQuad facilitates assessment of existing or planned mobile radio systems and wireless phones attached through various

[3]PESQ is a full-reference signal-layer speech quality assessment algorithm described in ITU-T Rec. P.862.

[4]www.gl.com

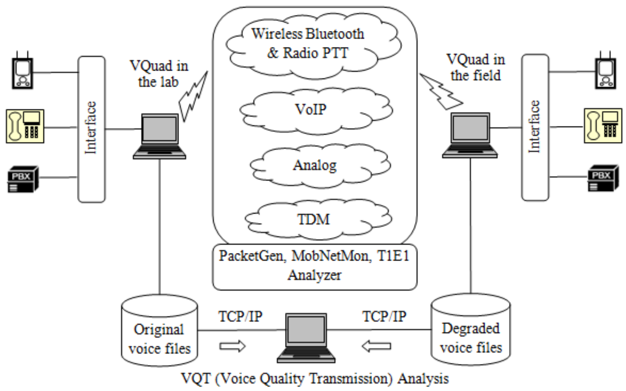


Fig. 5. Set-up for evaluating speech quality using VQuad [38].

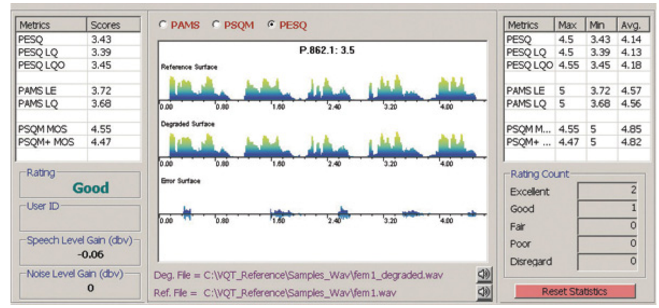


Fig. 6. Illustrative results of the output of VQuad.

kinds of access technologies, such as Wi-Max, Wi-Fi, and Bluetooth. VQuad quantifies the speech quality using the assessment algorithm defined in ITU-T Rec. P.862, PAMS (Perceptual Analysis Measurement System), and PSQM (Perceptual Speech Quality Measure). As such, it follows the operational mode of intrusive SQA algorithms. Notice that VQuad can be configured according to intended purposes either as a *real-time stand-alone hardware tester* or as a *software suite running on a calculator*. Figure 5 depicts a generic network set-up required to assess transport system and terminals using the intrusive SQA software VQuad. Notice that the GL Corporation has developed several simulators of network impairments and a wide variety of physical links.

From a practical viewpoint, the VQuad application should be installed at ingress and egress VoIP gateways. A global monitoring server should be set-up and properly configured to communicate with dispersed VQuad clients. As we can see in Figure 5, a reliable TCP/IP connection is set-up between the monitor and each VQuad client that is used for the delivery of reference and degraded speech contents. Once the reference and degraded speech sequences are downloaded, a full-reference SQA algorithm is invoked to measure the speech quality at the monitoring device. The obtained score is recorded and may be sent to the sender or the VoIP service provider as a feedback for a possible adaptation.

The screen snapshot given in Figure 6 gives representative results obtained using VQuad. The degraded and reference speech sequences can be played for a quick subjective verification. As we can note, multiple full-reference standardized SQA algorithms and versions can be used for the evaluation of speech quality. The main window gives an idea about signal content of reference and degraded speech sequence as well as the corresponding difference as function of time. Note that VQuad offers useful tools for comprehensive statistical analysis of collected scores.

Similar to GL Corporation, OPTICOM[5] has developed a *proprietary* software suite, which can be run on an ordinary computer, called OPERA [39]. It enables assessing, analyzing, and reporting the speech or audio quality in a user friendly fashion. OPERA supports a wide variety of coding schemes and includes at the core several double-ended signal-layer SQA algorithms standardized by the ITU-T, namely P.861 and P.862. As such, OPERA operates as a double-sided (or intrusive) SQA algorithm. OPERA provides a set of graphical tools for speech wave processing to examine the effect of signal-related impairments in temporal, frequency and acoustic domains, such as background noise, talker and listener echoes, double talk, speech clipping, saturation level, signal attenuation, and artefacts caused by Voice Activity Detector (VAD) algorithms. In addition, also as in VQuad, OPERA offers stand-alone hardware and computer-based testing / assessing modes. In [40], A. E. Conway et al. extend OPERA to properly subsume impairments sustained by packet-based voice streams delivered over VoIP networks, such as independent and burst packet losses, packets' reordering, and delay jitter. To do that, the reference streams of voice packets are distorted using analytical models of sources of impairments calibrated using suitable parameters. For instance, Bernoulli and Gilbert models have been used to introduce temporally-independent and bursty packet losses in a controlled fashion. In addition, to introduce delay jitter, authors utilized a FIFO queuing model, which is shared between a periodic voice packet stream and background traffic that follow a Poisson distribution. This extension enabled important findings about the relationship between speech quality and sources of impairments observed over VoIP networks. More elaborate and comprehensive work has been conducted by J. Turunen et al. to quantify VoIP quality under several application settings, such as anti-jitter buffer configuration, coding scheme, PLC and FEC strategies [41]. Note that besides OPERA, OPTICOM developed a *proprietary* SQA software suit denoted as 3SQM that operates following *signal-layer single-ended methodology*. It includes at the core a customized implementation of the generic signal-layer no-reference SQA algorithm described in ITU-T Rec. P.563.

Under the scope of his doctoral study, C. Hoene has developed an open-source client/server based SQA software called Mongolia [33]. Mongolia offers a graphical user interface that enables evaluating the speech quality under a packet loss process which selectively drops speech packets according to their perceptual importance. The ultimate goal was to calculate the perceptual importance of each voice frame with respect to the overall perceived quality. This enables an efficient scheduling of packet delivery at sender and intermediate nodes under an arbitrary network condition. Moreover, the perceptual importance of speech frames may be used by the de-jittering policy to efficiently schedule the playback

[5] www.opticom.de

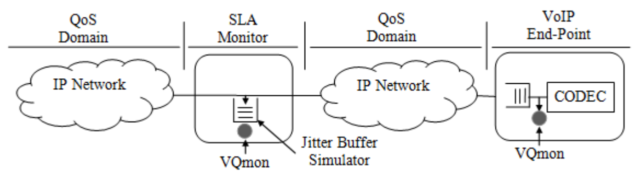


Fig. 7. Integration scenarios of VQmon in an existing IP-based network.

process. The developed framework Mongolia supports several popular low bit rate speech CODECs, namely ITU-T G.729 and GSM. Similar to OPERA and VQuad, Mongolia uses the standard SQA algorithm ITU-T P.862 to automatically calculate the perceived quality. As such, Mongolia falls into intrusive comparison-based SQA methodology.

Following the same tendency, A. Clark developed a *proprietary* SQA framework called VQmon that offers to users an easy-to-use graphical interface that enables the prediction of the perceived speech quality over an arbitrary IP configuration [42]. Its internal operational mode follows the *packet-layer single-ended parametric-model* SQA methodology. Hence, it is convenient for the sake of surveillance, live management, and troubleshooting of live VoIP conversations. Thus, it should be integrated at a SLA (Service Level Agreement) monitor or a VoIP edge (See Figure 7). The estimation of perceived quality of VoIP conversations using VQmon technology will be detailed in Section 4.2.1. VQmon allows users to accurately specify time-varying IP network impairments affecting the original packet stream at access and core stages.

The snapshot given in Figure 8 graphically illustrates the utilisation of VQmon for the evaluation of a simulated VoIP configuration. As we can see, a large number of network setting parameters can be assigned. Precisely, VQmon presumes that the access network is a LAN or a residential access segment that explains why packet ordering is preserved during this stage. Moreover, VQmon supposes that the core network is a collection of routers connected through high capacity links. In such a case, the sources of impairments observed in core IP-based networks are route switching and link failure that entail packet reordering, loss, and delay jitter. The framework illustrated in Figure 8 supports several CODECs, namely G.711 and G.729. The buffer size is maintained static during the whole VoIP session. As a result, the developed framework by Telchemy Corporation artificially synthesizes one-way network delays and packet loss. Moreover, it considers discarded voice packets at the de-jittering buffer. This information is used by VQmon to calculate the instantaneous rating factor as function of time (see Figure 8). At the end of the simulated call, VQmon calculates at call quality at MOS and R scales from listening and conversational perspectives.

Recently, S. Jelassi el al have developed an open-source software-based SQA framework of VoIP conversations delivered through a mobile ad-hoc network, known as EVOM [43]. For the sake of convenience, a graphical user interface has been set-up that enables defining network features, such as number of nodes, network dimension, mobility model, and routing protocol (see Figure 9). The developed SQA framework offers a flexible way to define network workloads

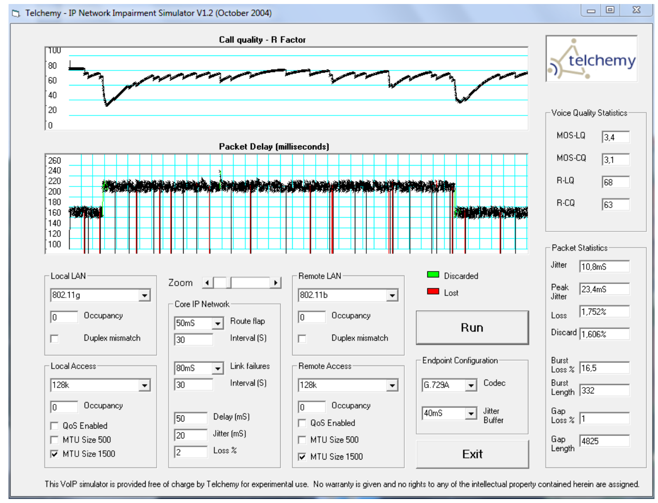


Fig. 8. Illustrative results given by VQmon using simulation software-suit developed Telchemy.

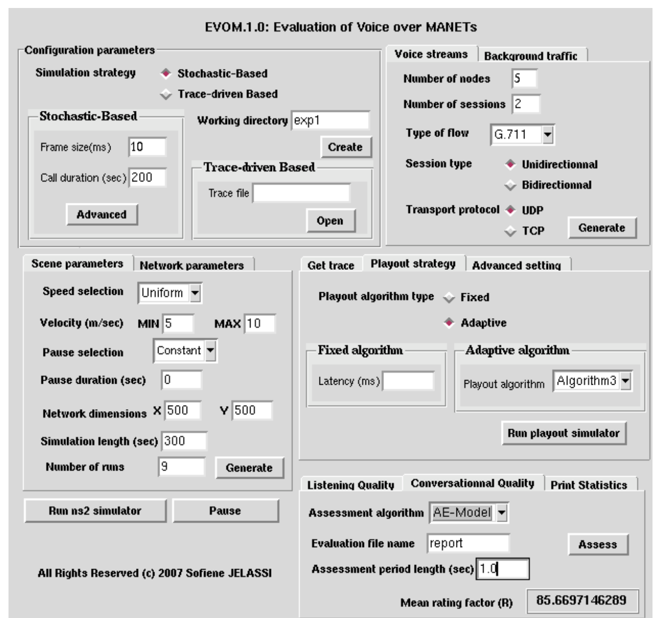
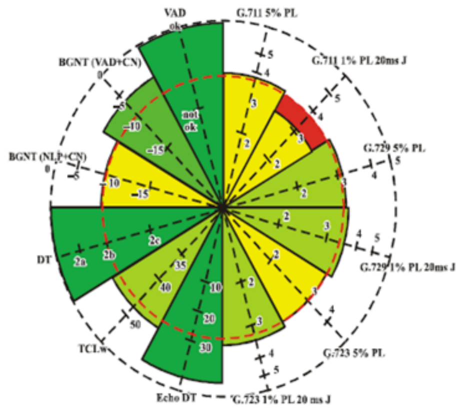


Fig. 9. Software-based framework for the evaluation of voice over a MANET.

consisting in a set of VoIP sessions and background traffic. The users have the ability to control the start time and duration of set-up VoIP sessions. Furthermore, the developed SQA framework supports several playback algorithms and coding schemes. Notice that the SQA framework in [43] is able to run simulation where injected data are generated according to a ON/OFF model or directly read from a pre-recorded trace. The SQA is made at packet (resp. signal) layer for stochastic (resp. trace) -driven simulation using a primitive parametric (resp. the ITU-T P.862) SQA algorithm.

Typically, a quality measurement procedure of a VoIP device (e.g. a cell-phone or a VoIP gateway) generates a lot of characterisation measures that reflects the implementation performance of included elements, such as CODEC, de-jittering algorithm, VAD, echo canceller, and noise removal [44]. For an ease- and quick- understanding, the ITU-T Rec. P.505 specifies a graphical presentation that exposes using



Interpretation

– The listening MOS quality of the cell-phone is given in the right halve of the quality pie gives. The red section under G.711 and mean jitter equal to 20 ms means that the cell-phone produces an unsatisfactory quality.

– The activation of echo suppression (NLP) and Comfort Noise (CN) leads to an important alteration of background noise (BGNT) features.

– The Voice Activity Detector (VAD) and CN are harmless of transmitted background noises (BGNT).

– The value of DT (Double talk) indicates that the cell-phone enables a full-duplex communication.

– The echo attenuation under single (TCLw) and double talk (DT) situations exceed the standard values.

Fig. 10. OVV (One View Visualization) of the performance of a cell-phone [45].

a *one-view visualization* (OVV) a plenty of characterization measures of tested VoIP devices [45]. The OVV diagram allows a quick identification of the strengths and weaknesses of a given black-box VoIP device. As such, a simple comparison between diverse cellular phones and VoIP gateways under a large variety of conditions and settings can be performed. The OVV plot has the form of a pie diagram divided into slices of equal area. An example of quality pie diagram and its correspondent interpretation is given in Figure 10.

The number of slices of an OVV diagram is equal to the cardinality of examined simple and compound quality-affecting parameters. As such, the spanning angle of each slice is equal to the ratio of 360 to the number of treated quality-affecting parameters. The axe scale of each slice is adjusted following the range values of metrics used to quantify quality-affecting parameters. The inner dashed boundary circle has a distinguished meaning following the specificity of each examined parameter. It may refer to the deviation of a given metric with respect to the average measured values (e.g., G.711 under packet loss and jitter) or recommended values (e.g., background noise and echo attenuation). The violation of such a threshold is visually emphasized using suitable color selection. Notice that T. Le Gall develops a primitive open-source Matlab toolbox for creating in a user friendly fashion such OVV pie diagrams of a given setting [46].

Before detailing single-ended parametric-model SQA algorithms in the next sections, we give in Figure 11 a qualitative clustering of SQA approaches. This plot enables to clarify research trends and to precisely situate our domain of exploration. As we can see, at the inception of quality-perception arena, laboratory-based subjective trials have been followed to understand the human auditory system. After the progression of speech processing systems, the first signal-layer SQA algorithms have been proposed and developed. They have been conceived for the automatic evaluation of speech quality processor, such as telephone devices, radio, and tape player. The evolving complexity of 1G telephone networks urged since early 80th the specification and development of parametric-model of telecom networks that are basically intended for planning purposes. After the blooming of digitized technology in the middle of 90th, the primary single-ended

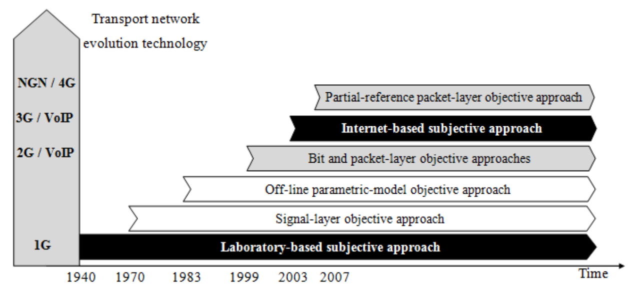


Fig. 11. A chronological positioning of existing SQA strategies.

parametric-model objective SQA approaches operating at bit- and packet- layers have been proposed and developed. Their goal is to evaluate voice quality over 2G (mainly GSM) and emerging VoIP networks. At early years of the last decade, Internet-based subjective SQA approach has been proposed and practiced. A few years ago, there is the proposal for a new SQA class that uses partial-signal meta-data conjointly with packet layer measures.

IV. Single-ended Parametric-model QoE Assessment Algorithms for VoIP

The pertinent feature of single-ended parametric-model SQA algorithms resides in their capacity to quantify at *run-time* the QoE of VoIP service using a set of network and application measurements. They perform a mapping operation of a set of *objective key parameters*, such as delay jitter and packet loss ratio to a *perceptual domain*. The basic building blocks of a generic single-ended parametric SQA algorithm are shown in Figure 12. Basically, a set of raw input parameters are *measured* then eventually *transformed* to increase their correlation with the response variable, i.e., a subjective score. Notice that input parameters are basically calculated using control information included in the header of received voice packets. Temporal processing is sometimes required to measure particular parameters, such as PLR (Packet Loss Ratio) and POR (Packet Out-of-Order Ratio). A *combination rule* is applied to estimate the listening or conversational quality of a monitored VoIP call, termed as LQE (Listening

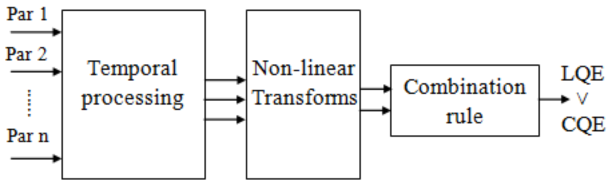


Fig. 12. Principle of single-ended parametric SQA algorithm [47].

Quality Estimate) and CQE (Conversational Quality Estimate), respectively.

The reported single-ended SQA algorithms can be discerned by the set of parameters used to characterize sustained distortions and the adopted *quality models* and *combination* rule that enable to capture the effect of various kinds of distortions. The ultimate goals are to provide a timely and accurate quality report feedback about an in-progress VoIP call. In the following sections, we detail several single-ended SQA algorithms recently proposed in the literature.

A. A primitive single-ended algorithm to evaluate VoIP service

R. G. Cole et al. developed the earliest prototype of a single-ended parametric SQA algorithm dedicated for live assessment of VoIP calls [48]. It constitutes a relevant adaptation of the ”baseline” offline parametric algorithm of the ITU-T E-Model. The alteration proposed by R. G. Cole et al. consists of using default values for E-Model parameters that are not related to the transport network, such as quantization, room and circuit noises. Moreover, they emphasize the effect of sources of quality degradation observed over data networks, namely one-way delay, packet loss ratio, and coding scheme. As such, the authors *periodically* calculate the rating factor using the following formula:

$$R = 94.2 - I_d(T_a) - I_e(CODEC, PLR) + A, \quad (1)$$

where, I_d and I_e are the *parametric models of distortions* that capture, respectively, the quality degradation caused by delay and equipment impairment factors. T_a is the mean one-way delay of played voice packets during an assessment interval, PLR is the packet loss ratio, and CODEC refers to the followed speech encoding scheme. The advantage factor A is used to quantify the users' willingness to accept quality degradation under certain situations, namely mobile communication.

To calculate R at run-time using Equation (1), adequate models of delay and equipment impairment factors should be developed. These models are built and calibrated following a regression process applied on *subjective results* given in ITU-T Rec. G.107. The following regressive model has been proposed to calculate the quality degradation caused by one-way delay when echoes are perfectly removed [48] as in Eq. (2).

Figure 13 graphically shows the variation of I_d as function of the one-way delay, T_a. As we can note, one-way delay less than 200 ms is harmless for the effectiveness of a voice

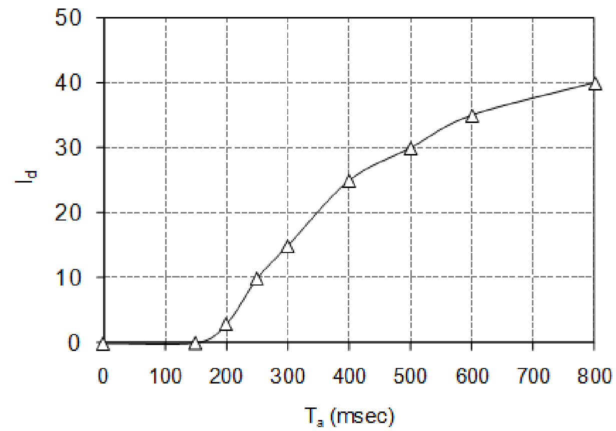


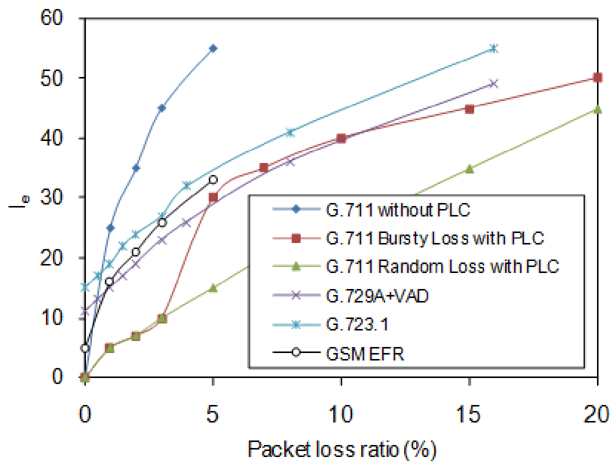
Fig. 13. Delay impairment factor, I_d, as function of mean network delay.

conversation. However, for one-way delay beyond 200 ms, the interaction quality of communicating entities is gradually decreased as the one way delay increases. In practice, voice transport configurations that produce a one-way delay higher than 400 ms should be avoided.

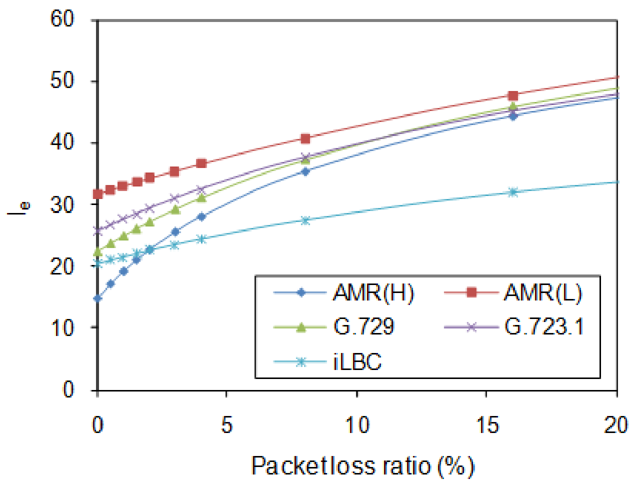
Recently, B. Sat et al. proposed new methodologies and measures to assess the effect of one-way delay on conversational quality [49]. In short, the authors analyze delays in the context of *face-to-face* and VoIP conversations. They defined the *human response delay* (HRD) or turn-taking latency as the silence duration between two successive active periods in a face-to-face voice conversation. In the context of VoIP conversations, the period separating the end of a spoken talk-spurt and the next listened one is referred to as the mutual silence (MS) that is tightly related to the M2E delay. Subsequently, author defined the *interactivity factor* (CI) of a given spoken and heard pair as the ratio of separating MS duration and the next HRD latency. Moreover, authors define the *conversational efficiency* (CE) as the ratio of the time a conversation takes in a face-to-face setting to the time to carry out the same conversation in a VoIP setting. CI and CE metrics are conjointly utilized to judge the conversational quality. Basically, the smaller (resp. greater) the value of CI (resp. CE) is, the better is the interactivity (resp. closeness to face to face conversation).

The perceived quality degradation captured by Ie is inherently linked to the speech coding scheme, packet loss behavior, de-jittering buffer policy, and packet loss concealment algorithm. In reality, existing speech coding schemes exhibit a distinct degree of resilience to deal with packet loss. This is illustrated in Figure 14 that shows the variation of the equipment impairment factor as function of packet loss ratio and coding scheme. As we can see, as a general rule, the higher the packet loss ratio, the lower the intelligibility of listened speech quality. However, the degrading amount caused by the packet loss process is linked to the features of each coding scheme. Note that curves given in Figure 14a (resp. 14b) has been obtained using the laboratory-based subjective (resp. signal-layer objective) ITU-T P.800 (resp. ITU-T P.862) approach.

$$
\begin{aligned}
I_d(T_a) &= 0.024 \times T_a + 0.11 \times (T_a - 177.3) \times H(T_a - 177.3), \\
\text{where } H(x) &= \begin{cases} 1 \; if \text{ x} < 0 \\ 0 \; if \text{ x} \geq 0 \end{cases}
\end{aligned} \tag{2}
$$



(a) Subjective I_e measures [50]



(b) PESQ-Based I_e measure [51]

Fig. 14. Equipment impairment factor, Ie, as function of packet loss ratio and coding scheme.

It is worth to note that emerging low bit rate predictive speech CODECs generate a set of dependent speech frames, which means that the faithful reconstruction of next frames requires the correct reception of previous ones. Note that the goal of designers of low bit rate predictive coding schemes is to reduce the bit rate without severely degrading the optimal perceived quality for narrowband spectrum obtained using the G.711 speech CODEC when PLC algorithm is activated [50]. The synchronization delay to generate again the original signal wave after a packet loss instance is dependent on the prediction scheme, the temporal location of missing slots, and the recovering strategy included in each speech CODEC. These features explain why the coding scheme and the characteristics of packet loss process are conjointly utilized as parameters of equipment impairment models. Based on subjective curve trends, R. Cole et al. suggest the following form for model I_e:

$$
I_e(CODEC, PLR) = a + b \times ln(1 + c \times PLR), \tag{3}
$$

where, a, b, and c are real fitting coefficients, obtained through a logarithmic regression analysis. For instance, the recommended coefficients for the G.711 speech CODEC when the packet loss concealment (PLC) algorithm is activated, under a random loss process, are a= 0, b= 30, and c= 15 [48]. To circumvent the need to conduct a subjective quality measurement campaign for each novel or upgraded speech CODEC, L. Sun et al. proposed an objective methodology that measures listening quality using the intrusive signal-layer SQA algorithm PESQ rather than human subjects [51]. Note that PESQ is well-recognized by its accuracy to evaluate speech quality on sequence-by-sequence basis under a wide range of conditions [16].

The basic single-ended SQA algorithm described in [48] is unable to account for several features of quality degradation sources sustained by voice calls over data IP networks, namely bursty packet loss behavior, unequal perceptual importance of speech signals carried in the payload, transient loss of connectivity, chain configuration, specific implementation of manufactured VoIP edge-devices, speech encoding scheme and rate changeover, de-jittering buffer dynamics, etc. Indeed, authors do not evaluate at all the correlation and accuracy of their produced CQE scores with CQS or CQO ones. This explains why several new, more sophisticated SQA algorithms dedicated for VoIP conversations have been reported in the literature, which will be detailed in following sections.

B. Augmented single-ended algorithms to evaluate VoIP service

In this section, we describe several new single-ended parametric SQA algorithms that try to improve the correlation between estimated and subjective quality scores over a wide range of network conditions and application settings.

1) Coarse- and subtle- grained SQA algorithms sensitive to packet loss burstiness: It is subjectively well-verified that the perceived quality of VoIP calls is tightly linked to the intensity of *packet loss burstiness*. This explains why packet loss processes distinctively affect the QoE of distorted speech sequence, in spite of having an equal mean PLR. Generally speaking, the higher the burstiness of packet loss process is, the lower is the users' satisfaction. To accurately estimate the QoE by end-users, the burstiness of packet loss process should be properly characterized and subsumed by SQA algorithm.

In this regard, A. D. Clark proposed a *coarse-grained* SQA algorithm, which constitutes the core of the SQA software suite VQmon described in Section 3, that accounts for a *limited range* of packet loss burstiness [42]. This signifies that the SQA algorithm is unable to properly discern between various distributions of bursty packet losses. The classification of loss instances either as isolated or bursty is performed through the monitoring of *inter-loss gap* and *loss run* durations. Precisely, a loss instance is considered as isolated if and only if its length is exactly equal to one speech frame and the duration since the former loss event is greater than gmin. The recommended value of gmin is equal to 320 ms that has been selected through subjective tuning process. This enables to subsume sparse loss distribution under high packet loss ratio. The *coarse-grained* features of packet loss process are captured using a 4-state Markov chain, rather than the classical 2-state Gilbert-Elliot model (see Figure 15a). The chain is calibrated at run-time using an efficient *packet loss event driven algorithm.* At the end of a monitoring period (8s - 20s), suitable parameters, such as average packet loss ratios for isolated and bursty packet loss periods and the correspondent durations are properly extracted. The value of Ie is separately calculated according to the average PLR under isolated and bursty packet loss periods, which are subsequently *integrated over time* to finally derive an *average value of* I_e. Notice that the integration subsumes users' behavior at the transition from isolated to bursty loss periods, and conversely. Moreover, the recency effect related to the temporal location of the last very poor quality period has been considered by VQmon to calculate the overall degradation at the end of a monitoring period. Notice that the VQmon assessment methodology has been used to fill extended RTCP reports described in the RFC 3611.

As indicated previously, VQmon is unable to subtly capture the perceived effect of a specific bursty packet loss pattern. For example, it is unable to discern between a situation where three packets are consecutively dropped and another where three packets are dropped with an inter-loss gap of 1 or more packets. These two loss patterns inherently lead to distinct listening qualities. To overcome this shortage, L. Roychoudhuri et al. proposed a *subtle-grained* SQA algorithm, denoted as Genome that more accurately subsumes the pattern of dropped speech frames [52]. To do that, a set of "base" quality estimate models which quantify the perceived quality entailed by the application of a periodic packet loss process[6] are developed, following a simple logarithmic regression analysis. The base quality estimate models are parameterized using the inter-loss gap and burst loss sizes. Specifically, for a packet loss run equal to 1, 2, 3, or 4 packets, a dedicated base quality estimate model, which has as input parameters the inter-loss gap size, has been built.

At run-time, Genome probes and records the effective experienced inter-loss gap and the following burst loss size. At the end of a monitoring period, the overall listening quality is computed as the weighted average of the "base" quality score of each pair, where the weights are calculated as a function of inter-loss gap durations (see Figure 15b). Notice that the combination formula of Genome implies that the larger the inter-loss gap size of a given pair, the greater the influence on the overall perceived quality. Moreover, a high frequency of a given pair entails more influence on the overall perceived quality. These statistical properties of Genome can result in a biased behavior rating. Moreover, the subtle granularity of Genome considerably disables its ability to subsume the context in which a given loss instance happens. This perhaps explains why authors confined the performance evaluation of Genome to randomly dropped speech packets.

Notice finally that an extended version of ITU-T E-Model has been released in 2005 that includes packet loss burstiness in the calculation of the equipment impairment factor, I_e [53]. In short, the designers of ITU-T E-Model introduce an additional parameter referred to as BurstR that is used to quantify the degree of packet loss burstiness and hence the amount of quality degradation. Moreover, an efficient window-based methodology, denoted as Q-Model, has been reported in the literature [54]. Q-Model maps a given bursty packet loss process to its equivalent random one with respect to quality degradation. As such, legacy quality models established for temporally-independent packet loss process can be reliably used to measure quality degradation.

In [55], S. Jelassi et al. compare the performance of previously described bursty-loss-aware SQA algorithms under a wide range of packet loss burstiness. The metrics used to judge the performance of examined SQA algorithms were the Pearson correlation coefficient and average absolute deviation between measured and estimated rating factors, denoted hereafter respectively as ρ and Δ. The value of Δ is obtained using the following expression:

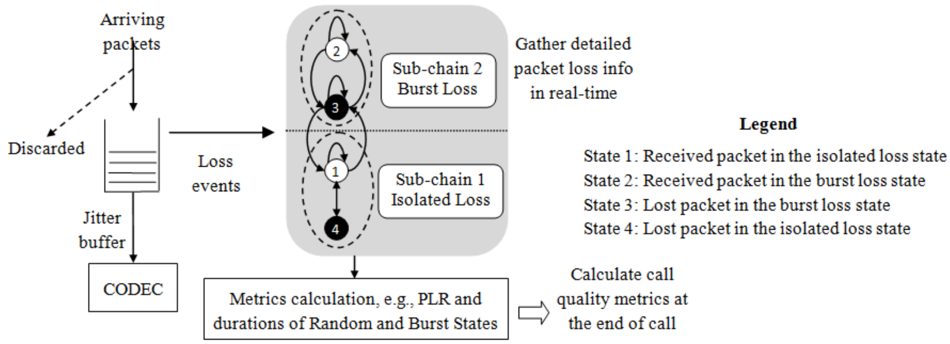
$$\Delta = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(R_M^i - R_E^i)^2} \tag{4}$$

where, R_M and R_E refer, respectively, to measured and estimated rating factors and N is the number of measures. The histograms given in Figure 16 give a recap of obtained results, where E-Model(1) and E-Model(2) denote, respectively, E-Model designed to subsume independent and bursty removed voice packets [53]. Q-Model(1) and Q-Model(2) refer, respectively, to two variants of Q-Model [54]. The value of ρ and Δ are calculated on sequence-by-sequence basis. We refer interested reader to [55] for further explanation about the empirical study and additional results.

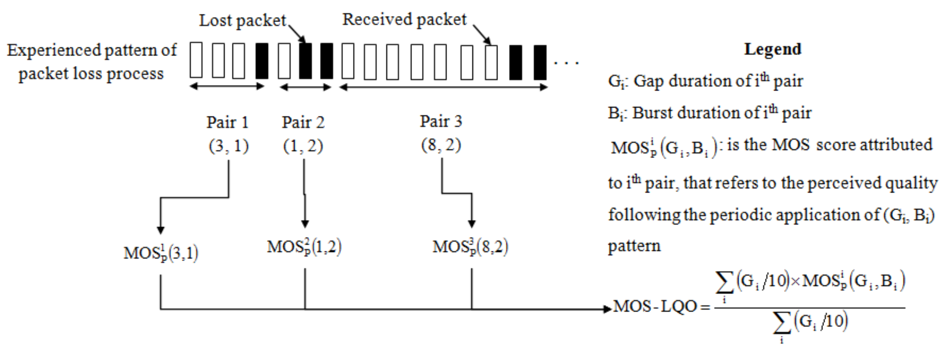
As we can see in Figure 16a, all bursty-loss-aware SQA achieve a good correlation coefficient greater than 0.85. This observation is somehow expected, as a significant increase of PLR values induces a considerable decrease of MOS scores, and conversely. All existing SQA algorithms are designed using *monotonic* quality models as function of PLR values, which explains the observed good correlation coefficients. As we can see, Q-Model(1) and Q-Model(2) slightly outperform other SQA approaches. Moreover, we see that VQmon achieves the minimum correlation coefficient following our measurements.

Histograms given in Fig. 16b summarize the obtained values of Δ of examined bursty-loss-aware SQA algorithms. As we can see, examined SQA algorithms induce significant deviation between measured and estimated scores. E-Model(1)

[6]A packet loss process that periodically drops a static number of consecutive speech frames preceded by a given inter-loss gap size.



(a) SQA methodology followed by VQmon [42]



(b) SQA methodology followed by Genome [52]

Fig. 15. SQA methodologies that account for packet loss burstiness in the calculation of final MOS score.

induces the maximal value of mean deviation, which is expected since it has been designed for independently removed packets. As we can see, Q-Model(2) achieves the minimum average deviation. The accuracy of E-Model(2) is better than E-Model(1) since it subsumes more properly packet loss burstiness. As we can note, the minimum values of Δ is roughly equal to 6, which in our opinion still pretty important. This constitutes the principal weakness and limitation of treated SQA, which should be tackled in future work.

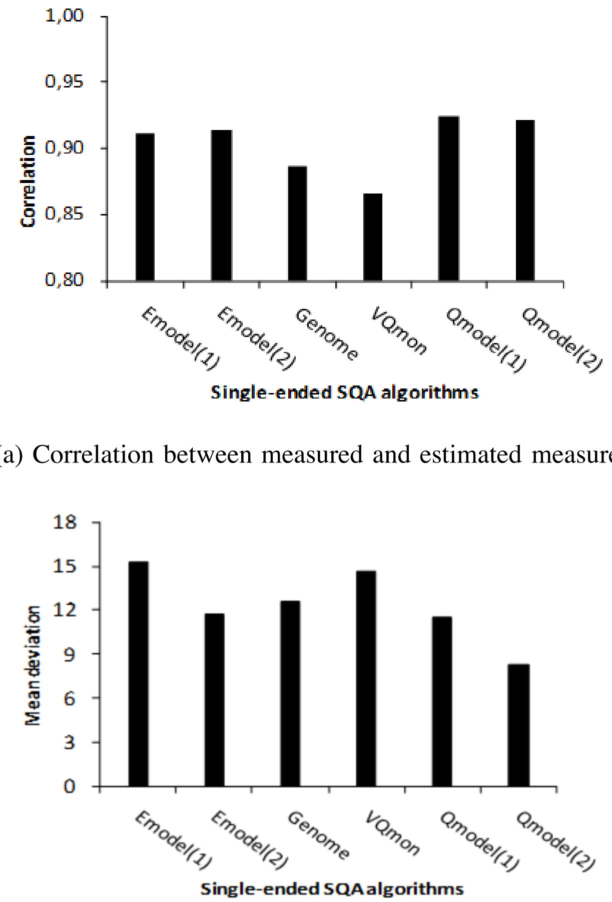
2) Edge-Aware SQA algorithms: A typical system configuration to deliver VoIP conversations is shown in Figure 17. As we can see, an edge-device is deployed at the end of each VoIP connection that can be either PC-to-PC or PC-to-PSTN voice conversation. The mechanisms to deal with network impairments integrated in each edge-device are specific to each manufacturer, and often undisclosed. The QoE is however tightly dependent on the effectiveness and efficiency of manufactured edge-devices. In fact, conducted empirical studies proved that diverse manufactured edge-devices produce very different QoE under a similar degree of network distortions, namely packet loss process and delay jitter [56], [57]. Notice that the previously described SQA algorithms VQmon and Genome assume that edge-devices are well-characterized, which is unrealistic in practice.

To subsume the features of a given manufactured edge-device, S. R. Broom proposed an edge-aware packet-layer SQA algorithm for VoIP conversations. In short, the conceived SQA algorithm monitors and updates a set of "base" parameters at the reception of each new voice packet, such as mean packet loss ratio, packet delay variation, loss and inter-loss durations. The "base" parameters are transformed then combined using *tailored* formulas and weighting coefficients specific to each edge-device profile that has been derived offline (before the service) through a comprehensive calibration and regression processes [56]. Formally, the following generic model has been used to estimate the QoE at the reception of nth packet:

$$Q^n = Q^{max} - H(\sum_{m=1}^{p} F_m X_{m,n}), \qquad (5)$$

where, $X_{m,n}$ refers to the value of mth base parameter at the reception of the nth packet, F_m is a non-linear function specifically developed for each edge-device, P is the number of parameters, H is a *monotonic* and *nonlinear* mapping function. Notice that the value returned by the function H represents the *overall quality degradation* on the MOS scale. This explains why the objective quality score, Q_n, is calculated as the maximum achievable MOS, Q_{MAX}, minus the overall quality degradation.

3) Artificial neural network-based single-ended SQA algorithms: For that sake of quality prediction, an Artificial Neural Network (ANN) can be utilized to emulate human behavior rating under a given condition [58], [59]. An overview for the development of ANN-based SQA algorithm is given in Figure 18. The preliminary phase consists of a subtle definition of sources of degradation over the transport system that will be utilized to carry live voice conversations. A representative speech database that includes a large variety of speech sequences should be methodically built. A selection of parameters that sensibly affect the perceived quality should



(a) Correlation between measured and estimated measures.

(b) Mean deviation between measured and estimated measures.

Fig. 16. Equipment impairment factor, Ie, as function of packet loss ratio and coding scheme.

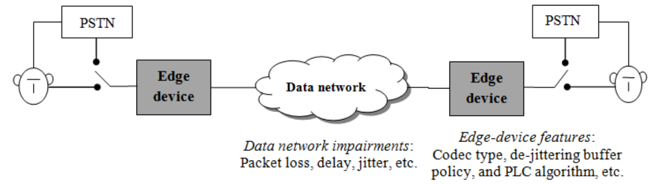


Fig. 17. Typical configuration of PC-to-PC or PC-to-PSTN connections [56].

TABLE I
STANDARD RELATIONSHIP BETWEEN THE RATING FACTOR AND TRANSMISSION QUALITY

R	Speech transmission quality
90 – 94.5	Excellent
80 – 90	High
70 – 80	Medium
60 – 70	Low
50 – 60	Poor
00 – 50	

be performed. Note that the selected parameters should be measurable at run-time using information available at the receiver. Then, range values should be associated with each selected parameter.

The original speech sequences are injected through the system under test. At its output, the degraded speech sequences are recorded. Moreover, a set of associate characterization measures are acquired from the system under test. The assessment of perceived quality of degraded speech sequences can be performed either using human subjects or an intrusive signal-layer SQA algorithm. Next, suitable neural network architecture should be defined. In [58] and [59], a three-layered feedforward network has been used. Certainly, the recorded values of the selected parameters will be the input to the ANN and the corresponding quality will be the output. The built database should be divided into two sets. The first one is used to calibrate the ANN and the second one is used to validate its accuracy. Once a stable neural network configuration is obtained, the suitable ANN's architecture and weights can be extracted in order to build a concise SQA tool. At run-time, the developed SQA algorithm gathers the values of parameters by monitoring received speech sequences. Once a new measure of quality is required, the calculated values are used by configured ANN to automatically derive MOS quality score.

The difficulty to manipulate ANNs explains why their application to evaluate VoIP is limited in practice. Indeed, previously described regressive quality models are simpler to derive and produces a comparable accuracy in term of correlation and precision. Moreover, the training phase is cumbersome and tightly related to a given configuration.

4) A SQA algorithm using the paradigm of equal perceptual quality contours: In contrast to classical telecom voice conversations where the perceived quality is roughly maintained static throughout the entire call, the perceived quality of a VoIP session is considerably variable. As such, a single score attributed at the end of a voice call is insufficient to study time-varying IP network impairments and the effectiveness of integrated mechanisms to deal with them. To better understand and investigate time-varying perceived speech quality, M. Narbutt et al. designed a SQA algorithm that constructs at the end of a VoIP conversation a pie diagram showing in a user friendly fashion the undergone perceived quality [60]. To do that, the authors utilize the relationship between the set combinations of quality-sensitive metrics, namely One-Way Delay (OWD) and PLR, the rating factor, and the speech transmission quality (STQ) to build the *equal perceptual quality contours* of a given VoIP configuration. Table I shows the relationship between the value of R and STQ [61].

Figure 19a graphically exhibits the paradigm of equal perceived quality contours of a VoIP configuration where echo is perfectly removed, speech wave is coded using G.711, and the PLC algorithm is activated. A given perceived region is characterized by an identical user satisfaction that is obtained for a given set of combinations of OWD and PLR. Notice that STQ regions are built offline using the primitive SQA algorithm described in Section 4.1 where the parametric models of impairment factors, I_e and I_d, are selected according to set-up VoIP configuration.

The authors divide the entire call period into small time intervals of duration 10 - 20 sec. The SQA algorithm measures the mean PLR and OWD metrics of each interval using recorded packet traces, then, it places the resulting combina-

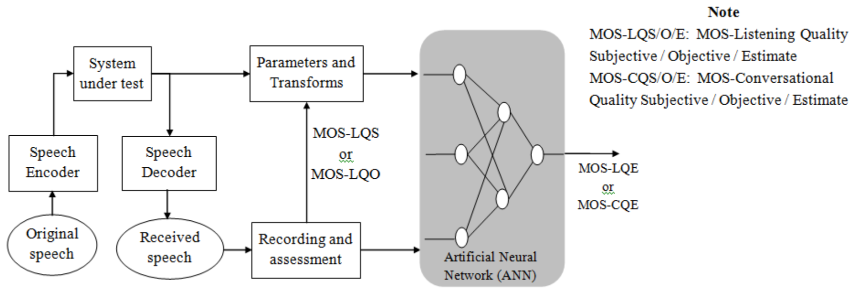
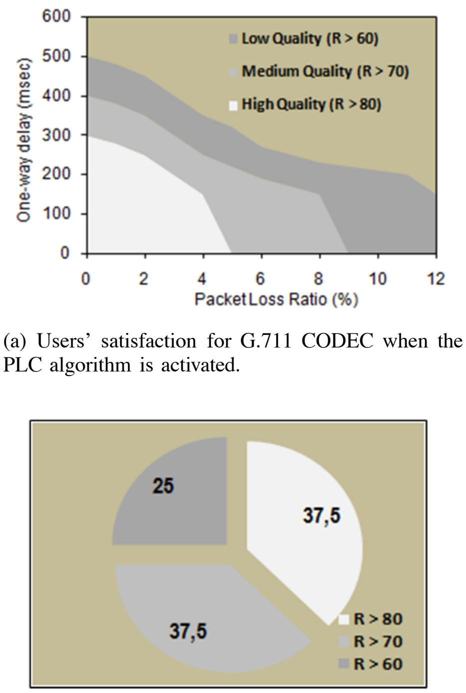


Fig. 18. Development of ANN-based SQA algorithm.



(a) Users' satisfaction for G.711 CODEC when the PLC algorithm is activated.

(b) An example of pie diagram presenting undergone STQ.

Fig. 19. Equal perceptual contours and their application for SQA.

tion on the built equal perceptual quality contours diagram. At the end of a VoIP conversation, the recorded history of STQ of each interval is presented in the form of a pie diagram. For the sake of illustration, we give in Figure 19b an example of a quality pie diagram, which indicates that the user perceives the quality of 37.5%, 37.5%, and 25% of the entire call duration respectively as high, medium, and bad qualities. As we can note, a pie chart presentation gives more insights about perceived quality throughout a VoIP session. Notice that this methodology has been effectively used by M. Narbutt et al. to precisely assess and explore the performance of several playback policies.

5) Parametric SQA models sensitive to perceptual importance of missing packets: The previously described SQA algorithms assume that all dropped voice packets have an *equal* effect on the perceived quality. However, it has been empirically shown that speech signals have an *unequal perceptual importance* [33], [62]. The goal of previous work regarding unequal perceptual importance of speech waves consists of studying and measuring the perceived effect of a missing speech frame that depends on the speech wave features carried in the media packet payload and on its temporal location. This specific property is unaccounted for by previously described single-ended SQA algorithms. To overcome this shortcoming, a new set of parametric speech quality models, which are sensitive to *packet loss burstiness* and *perceptual importance of missing packets*, has been proposed in [37]. The perceptual importance of a speech wave chunk depends tightly on its *voicing property*. Indeed, speech signals can be classified into *voiced sounds*, such as 'a' and 'o', *unvoiced sounds*, such as 'h' and 'sh', and *silence*. It has been empirically observed that loss of voiced sounds impair more severely the perceived quality than unvoiced ones.

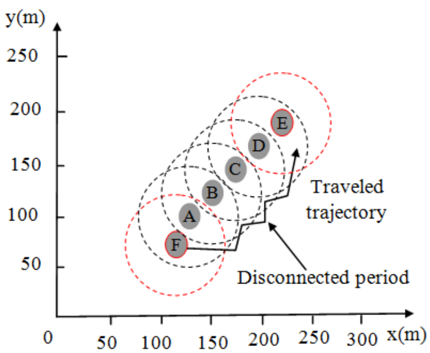
Following comprehensive empirical testing and statistical analyses, Jelassi el al. developed speech quality models which are able to quantify the effect of bursty packet loss processes that affect either voiced or unvoiced speech waves [63]. Authors select the PLR (Packet Loss Ratio) and the set of loss run durations for the sake of characterizing the packet loss burstiness. These parameters are transformed using tailored polynomial functions to increase their correlation with subjective scores. The produced voiced and unvoiced scores are combined to derive the overall score, which quantifies the perceived quality of removed voiced and unvoiced media packets. The receiver entity is notified about the voicing feature of speech waves included in dropped voice packets through a sender-based notification scheme.

The performance evaluation of voicing -aware and -oblivious speech quality estimate models for G.711 and G.729 conducted in [63] are summarized in Table II. The judgment metrics are the Pearson correlation and mean absolute precision. As we can note, the voicing-aware speech quality estimate models achieve a better correlation factor above 0.95 for both considered speech CODECs that is pretty satisfactory. Moreover, voicing-aware speech quality estimate models reduce notably, compared to voicing-oblivious speech quality estimate models, the mean absolute deviation between measured MOS-LQO and predicted MOS scores using voicing-aware models for both speech CODECs. The achieved accuracy is in the order of 0.2, which constitutes an excellent precision.

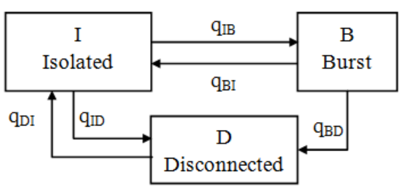
6) A disconnection-aware SQA algorithm: Considering their widespread deployment, it is desirable to integrate VoIPoW (VoIP over Wireless) to accommodate mobile users' needs, such as service ubiquity and flexibility. In addition to

TABLE II
PERFORMANCE COMPARISON BETWEEN VOICING AWARE AND OBLIVIOUS SPEECH QUALITY ESTIMATE MODELS

	Voicing-Oblivious Models [25]		**Voicing-Aware Models**	
	G.711	**G.729**	**G.711**	**G.729**
Correlation	0.927	0.910	0.954	0.961
Absolute mean deviation	0.61	0.92	0.22	0.17



(a) An episodic voice conversation over a MANET.



(b) Markov process model of packet losses over a wireless data channel.

Fig. 20. Transient loss of connectivity scenario and wireless channel modelling.

possible increased MAC contention delays, users may also suffer a *transient loss of connectivity*, which stems from inter- and intra- network handovers, or when users roam out of the coverage area of the associated infrastructure. Transient loss of connectivity is emphasized over Mobile Ad-hoc NETworks (MANET) characterized by their *episodic connections* caused by mobility of terminal as well as by forwarding nodes. A likely scenario is given in Figure 20a, where five nodes A, B, C, D, and E are retained stable and only node F travels with a fixed speed according to a zigzag trajectory [64]. Assume that a *voice session* is set-up between nodes E and F. In such a scenario, node F will undergo intermittent loss of connectivity with node E, and vice versa. This scenario illustrates the requirement to properly account for intermittent loss of connectivity on the QoE by designed SQA algorithms.

In this regard, S. Jelassi et al. propose a new single-ended parametric SQA algorithm designed to evaluate VoIP calls over *vanishing* wireless links, denoted as NIDA (Non-Intrusive Disconnection Aware assessment algorithm) [47]. It distinguishes between the perceptual effects of a single random packet loss, 2-4 consecutive packet losses (burst) stemming from contentions, and discontinuity entailed by transient loss of connectivity. To achieve this goal, a new wireless channel model has been developed following a continuous-time Markov process (see Figure 20b). The stochastic process has 3 states: I (for isolated), B (for burst), and D (for disconnected). When in state I, packet losses happen independently according to a Bernoulli distribution. However, when in state B, packet losses happen in bursts following the classical 2-state Gilbert model. When in state D, all sent packets are lost. The wireless channel model is calibrated at run-time using a set of measurements gathered from the header content of received media packets. The QoE under I, B, and D states is separately quantified at the end of a monitoring period using a set of parametric speech quality models [47]. The calculated quality measurements under each state are adequately combined to predict the quality degradation entailed by distortions observed over wireless channels.

For the sake of clarity and illustration, we summarize in Table III pertinent features of up-to-date and reviewed single-ended SQA algorithms of VoIP calls. As we can note, each single-ended in-service QoE assessment strategy of VoIP service subsumes only some factors that influence the QoE of VoIP service. The processing complexity and latency of parametric-model SQA algorithms are basically low because they only rely on some packet-layer measures to estimate the perceived quality. The wave signal itself is unconsidered. The raw resulting scores of each single-ended parametric-model SQA algorithm is given in the first column. Note that there is a possibility of mapping output SQA algorithms using standardized functions.

There are some flagship attempts to subsume quality-affecting factors and the operational mode of VoIP over emerging wireless networks, unconsidered by reported single-ended SQA algorithms, will be detailed in the next section.

V. PERCEIVED SPEECH QUALITY OVER EMERGING MOBILE AND PERVASIVE NETWORKS

The integration of VoIP over emerging mobile wireless data systems, namely Bluetooth, IEEE 802.16 and mesh networks is a key requirement for telecom stakeholders. In such a situation, the principal issue resides in the quick drop of the throughput as the number of hops and nodes increases [65]. This stems mainly from the access strategy followed by wireless MAC protocol to deliver a data packet [65]. In fact, one transmission needs the exchange RTS, CTS, and ACK control messages that hold the channel for a considerable duration. This overhead is greater for small packets, where the duration of control message is by far higher that the duration of sent useful data. Moreover, over multi-hop wireless connection, packets that belong to a given stream interfere to access to the wireless medium while crossing consecutive forwarding nodes. In [66], authors indicates that the capacity of a static string topology is reduced to one call when the number of wireless hops is equal to six.

There are several proposals for the improvement of wireless network capacity in term of supported number of calls. For

TABLE III
COMPARISON BETWEEN REVIEWED SINGLE-ENDED SQA ALGORITHMS OF VOIP CALLS

	Output scale	**Pertinent listening quality-affecting factors over VoIP transport systems**						**Indexes of performance**	
		Bursty losses	**Temporal alteration**	**Recency phenomenon**	**Specific edge implementation**	**Handover**	**CODEC changeover**	**Complexity**	**Latency**
VQmon[42]	R	CG	*	+	*	*	*	L	VL
Genome[52]	MOS	SG	*	*	*	*	*	L	VL
Edge-Aware[56]	MOS	SG	*	*	+	*	*	L	VL
ANN[59]	MOS	CG	*	*	*	*	*	M	VL
Perceptuel Contour[60]	Pie chart	*	*	+	*	*	*	L	VL
Voicing-Aware[63]	MOS	SG	*	*	*	*	*	L	VL
NIDA[47]	R	SG	*	+	*	+	*	L	VL

CG: Coarse Grained, SG: Subtle Grained
* : signifies a quality-affecting parameter that is unconsidered
+ : signifies a quality-affecting parameter that is considered
L: Low, VL : Very Low, M: Moderated

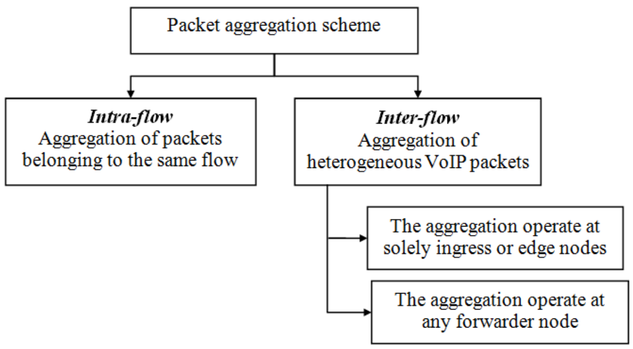


Fig. 21. Taxonomy of packet aggregation schemes over wireless networks.

instance, there is the proposal of using of multiple orthogonal frequencies to deliver voice packets [67]. This dramatically reduces signal inferences and increases network capacity. Moreover, there is the proposal of using cognitive frequency hopping policies that aim at smartly scheduling packet delivery among available frequencies [67]. The reduction of inference can be also performed using directional or multiple antennas. Apart increasing physical bandwidth of wireless links, there is a possibility to increase the number of simultaneous supported voice calls using packet aggregation [66], [68]. The idea behind such a strategy consists of delivering several small data packets residing in the waiting queue of an intermediate node during a single MAC transfer instance. Besides improving bandwidth efficiency, packet aggregation technique reduces considerably delay, jitter, and packet loss. The existing packet aggregation schemes have been summarized in Figure 21. Note that intra-flow packet aggregation scheme enabled to multiply the capacity of a string wireless topology by a factor of four [66].

A similar packet aggregation scheme has been proposed for last-hop IEEE 802.16 networks [69]. The authors propose to multiplex at a given VoIP gateway multiple inter-flow voice packets into a single larger one. The resulting packet is multicasted at the IEEE 802.16 AP toward connected wireless end devices. A de-multiplexer installed on each end device extracts its respective data and forwards them to the application. The performance evaluation strategy of packets' multiplexing scheme shows that it increases by a factor of two the network capacity in term of simultaneously supported calls [69]. A similar paradigm of packet aggregation methodology consists of concatenating an adaptive number of voice frames at edge-devices [70]. This end-to-end methodology optimizes delay and bandwidth tradeoff and improves application robustness over lossy wireless links [70].

Besides packets' aggregation, the enhancement of wireless channel utilization can be performed through the compression of packets' header [66], [68]. This was motivated by the fact that IP packets' header constitutes a considerable portion of delivered voice packets. Moreover, the IP packets' header is characterized by its redundancy. A generic packet header compression technique involves a compressor and decompressor deployed respectively at sender and receiver sides. The compressor and decompressor states should be synchronized, either implicitly (optimistic mode) or explicitly (acknowledged mode). ROHC and cRTP are well-known strategies for packets' header compression over connections including low link capacity. A special Zero-Length Header compression (ZLH) scheme has been proposed in [66]. ZLH is conjointly working with an aggregator of intra-flow VoIP packets. The idea behind ZLH consists of solely keeping the header of the first voice packet and removing the following packet headers included into the multiplexed packet. Given that included small packets are fully-ordered, the decompressor will be able to rebuild removed headers. The ZLH accompanied to intra-packet aggregation scheme increases the capacity of a static string wireless topology by a factor of nine [66].

Apart the issue of throughput over wireless networks, VoIP calls over *mobile and heterogeneous networks* will be obviously subject to handovers and roaming among multiple access networks [71]. A likely scenario for the provision of VoIP over a *heterogeneous* mobile network is illustrated in Figure 22. In such a situation, mobile devices will be likely equipped with multiple network interfaces, which mean that they are able to be attached to various categories of networks. A multitude of optimization strategies could be used focusing on users QoE, scarce resource utilization, and call price.

In such a context, a mobile customer can initiate, for example, a voice call using a *broadband wireless access network*, such as Wi-Fi or Wi-Max. The high network capacity enables using a *wideband* speech CODEC to encode the raw voice signal, characterized by its high data rate and excellent perceived quality. The attachment point of mobile users can de switched to *a narrowband access network* when the active mobile user moves outside the area covered by the broadband network or when the offered QoE is unsatisfactory[7]. In such a case, since network capacity is relatively reduced, the application should *changeover* the encoding technique from the wideband to narrowband speech encoding scheme. As such, the STQ is adapted following the prevailing link conditions

[7]In fact, over wireless data networks, the quality is time-varying and depends on the number of active users.

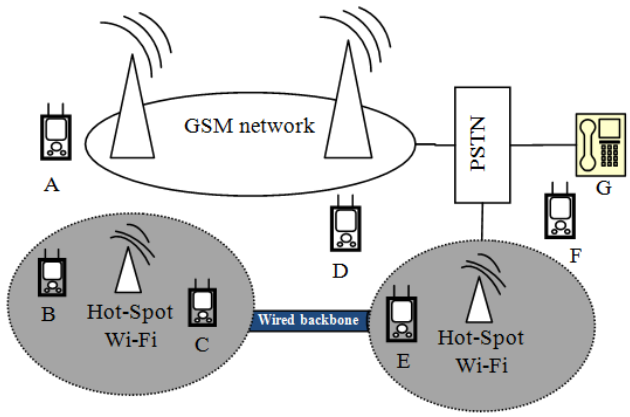


Fig. 22. VoIP service over heterogeneous networks.

that aim at the optimization of the users' QoE and network resource utilization.

In this regard, a research project, known as Mobisense, has been conducted at Deutsche Telekom lab located in Berlin [72]. The ultimate goal of Mobisense was to define subjective test methodologies that enable subtle quantification of QoE of VoIP conversations over mobile heterogeneous networks. A comprehensive subjective test companion has been conducted that reveals the following findings:

1) The packet loss process and audio bandwidth, i.e., wide- and narrow- bands, are the key factors that affect the QoE of a VoIP conversation.
2) Switching the audio bandwidth is roughly equivalent to the quality degradation of 5-10% packet loss, in both NB and WB conditions.
3) The degradations due to soft network handover (make-before-break) alone, i.e., without codec switching, are negligible in comparison to packet loss and audio bandwidth switching.
4) A changeover to a NB speech CODEC that reduces packet loss ratio is more advantageous than pursuing utilization of WB speech CODEC.
5) A switching from NB to WB speech transmission sensibly upgrades the perceived quality solely when remaining call duration is sufficiently long (30 sec). A switching from WB to NB is always linked to a loss in quality
6) The following relationship has been observed between packet-loss and audio-bandwidth degradations:
    - If packet loss is high (low basic quality), the impact of switching audio bandwidth on perceived quality is low.
    - If packet loss is low (high basic quality), the impact of switching audio bandwidth on overall call quality is high.
7) In bad network conditions (high packet loss ratio), it is not important whether the audio bandwidth can be maintained or not. The potential effort should concentrate on the reduction of packet loss. Hence, a handover that reduces packet loss ratio should be made in such a context.

The obtained findings can be used to design new management policies, and especially efficient and quality-aware handover strategies. In reality, handover can be triggered either in terminal- or network- centric fashions. Moreover, handover can involve relaying mobile nodes and randomly distributed proxy servers. The decision to execute a handover and followed strategy will be performed in such a way that it optimizes the QoE. Apart the mobility management, the QoE can be useful for other management operations, such as admission control, topology control, network selection, and service charging.

A multitude of questions remain that require more investigation in such a context. These include:

- Trans-coding operation that can be made over heterogonous networks.
- The perceived effects of alteration of background noise when a user moves among multiple environments during a conversation, such as a car, a street, and home.
- The perceived quality of a speech CODEC that generate multiple sub-streams, e.g., a base layer stream and several enhancement layer streams.
- The utilization of ultra-wide-band and high bit rate speech CODECs, e.g., MP3, MPEG, AAC.
- The perceived effect of aggregation and header compression schemes voice calls over multi-hop wireless networks.
- The perceptual effect of transient loss of connectivity.

VI. Open Issues Related to SQA of VoIP over NGN Systems

Despite the research summarized to date and issues listed above for heterogeneous mobile networks, a range of further issues and challenges remain that require attention in order to more accurately quantify users' QoE. In this section, we enumerate some potential issues.

A. *Perceived effect of buffer de-jittering policies*

VoIP receiver nodes remove the *network delay jitter* through a *buffering mechanism*, which aims to restore the original uniform interval separating successive media packets. In transit, media packets typically undergo variable one-way network delays that obliterate the temporal structure of original packet stream. To compensate for this, a dynamic de-jittering buffer *policy* seeks to predict the optimal *playback delay* that will optimise the trade-off between the *total one-way delay* and *late arrival ratio*. The playback delay is typically adjusted at the start of a new talk-spurt, which alters the original silence periods. Such a mechanism leads to the *expansion* (see Figure 23B) or *contraction* (see Figure 23C) of the original silence periods between utterances.

To date, the perceived disturbing effects caused by such alterations of the temporal structure in the original stream have been largely neglected [73]. However, anecdotal evidence and some preliminary research suggest that beyond a certain threshold, expansion and contraction will surely influence the users' QoE [74]. Indeed, the SQA algorithm defined in ITU-T Rec. P.862 that is recognized by its outstanding accuracy performs time alignment of the reference and degraded speech

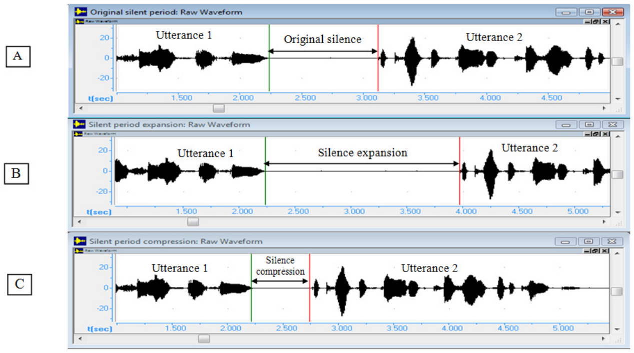


Fig. 23. Expansion and reduction of inter-utterance silence period duration.

signals before starting the assessment process. This signifies that temporal alterations of reference signal are completely ignored. That was acceptable for telecom transport systems, which introduce perceptually negligible temporal impairments. However, reference speech signal delivered over IP-based network are subject to temporal impairments with significant magnitude that lead to clearly perceived distortions. Hence, new SQA algorithms need to subsume such *temporal impairments* taking into account the *temporal location*, *frequency*, and *magnitude* of playback delay adjustments. Indeed, it is expected that the perceived effect of inter-word delay adjustments will be distinguished from inter-utterance delay adjustments. We believe that extensive subjective trials should be conducted to properly quantify and thus model the perceived effect of such temporal impairments.

Besides ignoring temporal impairments, the designers of the ITU-T PESQ algorithm indicated that it exhibited other shortcomings, especially for delivery systems that substitute speech wave with a silence [77]. This could happen at the front- and back- end of a given talk-spurt. Moreover, such an event can be observed when a missing packet is replaced with silence. Large deviations have been observed in cases where complete words or even sentences are removed from the reference speech signal. As such, ITU-T PESQ is unable to assess bursty packet loss behavior that leads to drop a large number of successive packets (¿ 420ms). We believe that such a feature should be included into ITU-T PESQ using subjective testing. Notice that despite this clear limitation, A. Radwan et al. utilized ITU-T PESQ to model the effect of temporal clipping[8] caused by echo canceller or voice activity detector algorithm [36]. This urges to review proposed and patented models for better QoE prediction, especially, when temporal clippings are frequent and persistent.

B. Subtle assessment of VoIP applications in practical conditions

To deal with a wide range of network conditions, VoIP applications should exhibit a high degree of resilience. As such, it should be able to adapt their packet transmission rate and scheduling, coding scheme, packetizing methodology, and packet protection, according to prevailing network condition.

[8]A clipping consists of substituting a noise or speech like noise signals by a silence

Moreover, many VoIP applications utilize a VAD (Voice Activity Detector) algorithm to discern between active and silence periods of a speech sequence. Typically, a bridging mechanism of successive active periods separated by a short silence one is utilized to prevent temporal clippings that considerably degrade the perceived quality. Furthermore, to generate a realistic silence rather than a mute wave signal, the description of ambient noise of the idle side is periodically transmitted toward the active side.

Typically, popular VoIP applications, such as Skype, Google-Talk, Yahoo Messenger and Windows Live keep disclosed their adaptive behavior to deal with time-varying impairments. In reality, considerable engineering effort is required to study and understand the running mode of popular VoIP applications [76], [78]. Nowadays, there is no uniform dedicated subjective test methodology for the assessment of perceptual performance of popular VoIP applications. Hence, we believe that there is an urgent requirement to develop a homogeneous subjective test methodology intended especially for the assessment of VoIP applications. It should be designed in such a way that it enables to study the perceived quality under realistic conditions from listening and conversational perspectives. Table IV summarises several test methodologies selected after a research bibliography work to assess the *conversational quality* of VoIP calls.

Basically, existing SQA algorithms of VoIP focus on the accurate estimate of *listening quality*. This choice can be explained by the expected considerable impact of equipment-related impairments, namely coding scheme and packet loss on the overall perceived quality. This means that perceived effects of delay-related impairments are neglected through the assumption that the perceived effect of one-way delay is common under all conversation tasks and scenarios, e.g., PC-to-PC and PC-to-PSTN. In fact, existing SQA algorithms basically utilize the primitive models given in Equation 2 to quantify the effect of one-way delay. In our opinion, there is a requirement to develop more elaborated quality models for the quantification of delay impairments that consider interactivity and seamlessness of a monitored VoIP call. In the context of the ITU-T E-Model, new models of impairment factor Id should be developed using suitable characterization parameters. Table III summarizes several metrics that can be used to judge the conversational quality of VoIP calls. As we can note, each approach utilizes a specific set of measures that are dedicated for VoIP conversation due to the lack of standards. This needs extensive subjective testing and an elaborated statistical analysis. Such an extension will enable to fairly judge the performance of popular VoIP applications under real conditions.

C. Perceptual importance of wideband speech CODEC frames

As indicated previously, earlier empirical subjective and objective studies have shown that certain speech segments have very different effects on perceived quality than others. In [33], C. Hoene develops a primitive methodology to quantify the perceptual importance of each speech frame in the context of a narrowband speech CODEC. Notice that the perceptual importance can be useful for smart transmission scheduling at

TABLE IV
COMPARISON BETWEEN UTILIZED SUBJECTIVE TEST METHODOLOGIES FOR THE ASSESSEMENT OF VOIP CALLS

	Approach of F. Hammer et al. [75]	**Approach of K-T Chen et al. [76]**	**Approach of B. Sat et al. [49]**
Conversation scenarios	Guided conversation tasks as specified in [17]	Empirical arbitrary gathered VoIP conversations	Simulation of pre-recorded Face to Face conversations
Methodology for impairment introduction	Simulation	Empirical	Emulation and Simulation
Assessment strategy	Objective	Objective	Subjective and Objective
Characterization metrics of conversational quality	SAR[1] and IR[2]	R[2], RD[3], and TL[4]	CI[5] and CE[6]

[1]SAR : Speaker Alternation Rate; the number of turn-taking between the talkers per minute
[2]IR : Interruption Rate; the number of times where a speaker is interrupted per minute
[3]R : Responsiveness; the proportion of OFF periods that coincide with ON periods of the other side
[4]RD : Response Delay; response delay after the opposite side stop talking
[5]TL : Talk-spurt Length; the duration of a talk-spurt
[6]CI : Coefficient of Interactivity; the ratio between two successive silences of a given side
[7]CE : Conversational Efficiency; the ratio of call duration in the context of face to face and VoIP

edge-device and relaying nodes. As such, an optimal perceived quality can be achieved for an arbitrary network condition.

This approach needs to be extended to emerging wide-band speech CODECs. Moreover, it is relevant to develop suitable methodologies to quantify frame importance in the context of a trans-coded speech signal. In such cases, the importance of each frame should be rigorously defined and characterized. Indeed, the importance of given speech frame is tightly dependent on the context and on the temporal location within the speech signal. We believe that an adaptive windowing mechanism will be helpful to accurately estimate the importance of a speech frame under given conditions.

Finally, notice that providing responses to such issues need a comprehensive investigation, which requires the specification of suitable realistic scenarios and extensive subjective testing. New quality models should be developed following a rigorous statistical methodology.

VII. CONCLUSIONS AND FUTURE WORK

This paper reviews in-depth the topic of speech quality assessment over time-varying QoS transport systems. It summarizes various subjective and objective methodologies defined in the literature for the assessment of perceived quality. In particular it focuses on single-ended SQA algorithms characterized by their attractive features useful for several applications, such as management, monitoring, diagnosis, dynamic configuration, and billing. It surveys single-ended SQA algorithms proposed recently in order to reduce the distance between objective and subjective quality measurements under a wide range of conditions. Among other issues, it examines the effect of time-varying QoS over data transport networks, the perceptual importance of presented speech signals, and the frequently observed transient losses of connectivity over wireless networks. Looking to the future, the need to subsume the perceived effect of an adaptive network delay jitter removal policy, which alters the temporal structure of original stream, in the estimation of overall perceived quality is outlined. Moreover, several new mechanisms, which affect the perceived quality, over heterogeneous mobile networks are summarized. Furthermore, this survey points out the fact that Multimodal and Multiparty conversations that involve voice and video streams and many collaborators should be addressed in detail to identify sources of quality degradation and to propose suitable remedies.

REFERENCES

[1] J. Hassan, S. K. Das, M. Hassan, C. Bisdikian, and D. Soldani, "Improving quality of experience for network services," *IEEE Network*, vol. 24, pp. 4–6, Mar. 2010.

[2] H. Koumaras, F. Liberal, and L. Sun, "Quality of experience issues in multimedia provision," *Telecommunication Systems*, pp. 1–3, Jun. 2010.

[3] P. Brooks and B. Hestnes, "User measures of quality of experience: why being objective and quantitative is important," *IEEE Network*, vol. 24, no. 2, pp. 8–13, Mar. 2010.

[4] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "Quadrant of euphoria: a crowdsourcing platform for qoe assessment," *IEEE Network*, vol. 24, no. 2, pp. 28–35, Mar. 2010.

[5] A. Raake, *Speech Quality of VoIP: Assessment and Prediction.* John Wiley & Sons, 2006.

[6] ITU-T. (2010) Itu-t study group 12 (study period 2009-2012). [Online]. Available: http://www.itu.int/ITU-T/studygroubs/com12/index.asp

[7] G. Scheets, M. Parperis, and R. Singh, "Voice over the internet: A tutorial discussing problems and solutions associated with alternative transport," *IEEE Commun. Surveys Tutorials*, vol. 6, no. 1-4, pp. 22–31, 2004.

[8] I. Telegeography. (2010, Mar.) International carriers' traffic grows despite skype popularity. [Online]. Available: www.telegeography.com

[9] J. Ellis and C. P. J. Rahman, *Voice, video, and data network convergence: architecture and design, from VoIP to wireless.* Academic Press, 2003.

[10] Y. Pan, "Digital audio compression," *Digital Technical Journal*, vol. 5, no. 2, pp. 1–14, 1993.

[11] H. Schulzrinne and E. Wedlund, "Application-layer mobility using sip," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 4, pp. 47–57, Jul. 2000.

[12] B. Li, M. Hamdi, D. Jiang, and X.-R. Cao, "Qos-enabled voice support in the next-generation internet: Issues, existing approaches, and challenges," *IEEE Commun. Mag.*, vol. 38, no. 4, pp. 54–61, Apr. 2000.

[13] V. Srivastava, , and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 112–119, Dec. 2005.

[14] L. Roychoudhuri, "Proactive rate and error control for packet multimedia transmissions based on loss prediction," Ph.D. dissertation, School of Computer Science, Telecommunications and Information Systems, DePaul University, May 2008.

[15] Z. Li, L. Sun, Z. Qiao, and E. Ifeachor, "Perceived speech quality driven retransmission mechanism for wireless voip," in *Proc. IEEE Fourth International Conference on 3G Mobile Communication Technologies*, London, UK, Jun. 25-27, 2003, pp. 395–399.

[16] A. W. Rix, J. G. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality - technology and applications," *IEEE Trans. Audio Speech Language Processing*, vol. 14, no. 6, pp. 1890–1901, 2006.

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

[18] S. Quackenbush, T. Burnwell, and M. Clements, *Objective Measures of Speech Quality.* Prentice Hall, 1988.

[19] *Continuous Evaluation of Time-Varying Speech Quality*, ITU-T Std. Recommendation P.880, 2005.

[20] *Study the Relationship between Instantaneous and Overall Subjective Speech Quality for Time-Varying Speech Sequence: Influence of a Recency Effect*, ITU-T Std. ITU Study Group 12, Contribution D.139 (France Telecom), 2000.

[21] *One-Way Transmission Delay*, ITU-T Std. Recommendation G.114, 2003.

[22] N. Kitawaki and K. Itoh, "Pure delay effects on speech quality in telecommunications," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 4, pp. 586–593, May 1991.

[23] L. Sun and E. Ifeachor, "New methods for voice quality evaluation for ip networks," in *Proc. 18-th International Teletraffic Congress (ITC18)*, Berlin, Germany, Aug.31 Sep 05 2003, pp. 395–399.

[24] K.-T. Chen, C.-C. Tu, and W.-C. Xiao, "Oneclick: A framework for measuring network quality of experience," in *Proc. of INFOCOM 2009*, Rio de Janeiro, Brazil, Apr. 19-25, 2009, pp. 702–710.

[25] L. Sun and E. C. Ifeachor, “Voice quality prediction models and their application in voip networks,” *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 809–820, 2006.

[26] S. Voran, “Objective estimation of perceived speech quality, part ii: Evaluation of the measuring normalizing block technique,” *IEEE Trans. Speech Audio Processing*, vol. 7, no. 4, pp. 383–390, Jul. 1999.

[27] N. Kitawaki, H. Nagabuchi, and K. Itoh, “Objective evaluation for low-bit-rate coding systems,” *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 242–248, Feb. 1988.

[28] S. Rein, F. H. P. Fitzek, and M. Reisslein, “Voice quality evaluation in wireless packet communication systems: a tutorial and perfomance results for rohc,” *IEEE Wireless Commun.*, vol. 12, no. 1, pp. 60–67, Feb. 2005.

[29] *Single-ended method for objective speech quality assessment in narrow-band telephony applications*, ITU-T Std. Recommendation P.563, 2004.

[30] *The E-Model, A Computational Model for Use in Transmission Planning*, ITU-T Std. Recommendation G.107, 2000.

[31] N. Kitawaki, “Perspectives on multimedia quality prediction methodologies for advanced mobile and ip-based telephony,” in *Proc. Workshop on Wideband Speech Quality in Terminals and Networks: Assessment and Prediction*, Mainz, Germany, Jun. 8-9, 2004.

[32] A. Takahashi, “Opinion model for estimating conversational quality of voip,” in *Proc. 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’04)*, Honolulu, Hawaii, USA, Apr. 15-20, 2004.

[33] C. Hoene, “Internet telephony over wireless links,” Ph.D. dissertation, Technical University of Berlin, Germany, Dec. 2005.

[34] S. Moller, A. Raake, N. Kitawaki, A. Takahashi, and M. Waltermann, “Impairment factor framework for wide-band speech codecs,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1969–1976, Nov. 2006.

[35] A. Takahashi, A. Kurashima, and H. Yoshino, “Objective quality methodology for estimating conversational quality in voip,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1984–1993, 2006.

[36] L. Ding, Z. Lin, A. Radwan, M. S. El-Hennawey, and R. A. Goubran, “Non-intrusive single-ended speech quality assessment in voip,” *Speech Communication, Elsevier*, vol. 49, no. 6, pp. 477–489, 2007.

[37] S. Jelassi, H. Youssef, C. Hoene, and G. Pujolle, “Voicing-aware parametric speech quality models over voip networks,” in *Proc. 2nd IEEE Global Information Infrastructure Symposium (GIIS 2009)*, Hammamet, Tunisia, Jun. 23-26, 2009.

[38] G. Corporation. (2010, Mar. 23) Vquad: Voice quality testing across all networks. [Online]. Available: http://www.gl.com/vquad.html

[39] A. Rix. (2010, Sep. 03) Personal homepage of dr. antony rix. [Online]. Available: http://www.rixcam.com/

[40] A. E. Conway and Y. Zhu, “A simulation-based methodology and tool for automating the modeling and analysis of voice-over-ip perceptual quality,” *Performance Evaluation*, vol. 54, no. 2, pp. 129–147, 2003.

[41] J. Turunen, P. Loula, and T. Lipping, “Assessment of objective voice quality over best-effort networks,” *Computer Networks*, vol. 28, no. 5, pp. 582–588, Mar. 2005.

[42] A. Clark, “Modeling the effects of burst packet loss and recency on subjective voice quality,” in *Proc. 2nd IP-Telephony Workshop (IPTel’2001)*, Columbia University, New York City, USA, Apr. 2-3, 2001.

[43] S. Jelassi and H. Youssef, “Evom: A software based platform for voice transmission and quality assessment over wireless ad-hoc networks,” in *Proc. 10th IEEE/ACM Annual International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2007)*, Chania, Crete Island, Greece, Oct. 22-26, 2007.

[44] F. Kettler, N. Rohrer, and S. Poschen, “Voip speech quality test event,” in *Proc. 5th ETSI Speech Quality Test Event*, Herzogenrath, Germany, Jan.Dec 2006.

[45] *One-View Visualization of Speech Quality Measurement Results*, ITU-T Std. Recommendation P.505, 2005.

[46] T. L. Gall. (2009, Mar. 10) One view visualization computer program. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/25299-itu-t-p505-one-view-visualization-diagrams-gui

[47] S. Jelassi, H. Youssef, L. Sun, and G. Pujolle, “Nida: A parametric vocal quality assessment algorithm over transient connections,” in *Proc. 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services (MMNS 2009)*, Venice, Italy, Oct. 26 - 30 2009.

[48] R. G. Cole and J. H. Rosenbluth, “Voice over ip performance monitoring,” *Computer Communication Review, ACM SIGCOMM*, vol. 31, no. 2, pp. 09–24, 2001.

[49] B. Sat and B. W. Wah, “Analyzing voice quality in popular voip applications,” *IEEE Multimedia Mag.*, vol. 16, no. 1, pp. 46–59, Jan./Mar. 2009.

[50] *Provisional Planning Values for the Equipment Impairment Factor Ie*, ITU-T Std. Recommendation G.113 - Appendix I, 2001.

[51] L. Sun and E. Ifeachor, “New models for perceived voice quality prediction new models for perceived voice quality prediction optimization for voip networks,” in *Proc. IEEE International Conference on Communications (ICC 2004)*, Paris, France, Jan.20-24 2004, pp. 1478–1483.

[52] L. Roychoudhuri and E. Al-Shaer, “Real-time audio quality evaluation for adaptive multimedia protocols,” in *Proc. Multimedia Networks and Services (MMNS 2005)*, Spain, Oct. 24 - 26 2005.

[53] A. Raake, “Short- and long-term packet loss behavior: Towards speech quality prediction for arbitrary loss distributions,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1957–1968, Nov. 2006.

[54] H. Zhang, L. Xie, J. Byun, P. Flynn, and C. Shim, “Packet loss burstiness and enhancement to the e-model,” in *Proc. The 6th IEEE International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Towson, Maryland, USA, May23-25 2005.

[55] S. Jelassi and G. Rubino, “A comparison study of speech quality assessment algorithms sensitive to packet loss burstiness,” in *Proc. IEEE Consumer Communications and Networking Conference*, Las Vegas, Nevada, USA, Jan.09-12 2011.

[56] S. R. Broom, “Voip quality assessment: Taking account of the edge-device,” *IEEE Trans. Audio, Speech, And Language Processing*, vol. 14, no. 6, pp. 1977–1983, Nov. 2006.

[57] M. Masuda and T. Hayashi, “Non-intrusive quality monitoring method of voip speech based on network performance metrics,” *IEICE Trans. Communication*, vol. E89-B, no. 2, pp. 304–312, Feb. 2006.

[58] S. Mohamed, F. C.-Prez, and H. Afifi, “Audio quality assessment in packet networks: an inter-subjective neural network model,” in *Proc. 15th International Conference on Information Networking (ICOIN’01)*, Beppu City, Oita, Japan, Jan. 31 Feb. 02 2001.

[59] G. Rubino, P. Tirilly, and M. Varela, “Evaluating users’ satisfaction in packet networks using random neural networks,” in *Proc. 16-th International Conference on Artificial Neural Networks (ICANN 2006)*, Athens, Greece, Sep.10-14 2006.

[60] M. Narbutt, A. Kelly, L. Murphy, and P. Perry, “Adaptive voip playout scheduling: Assessing user satisfaction,” *IEEE Internet Computing*, vol. 9, no. 4, pp. 28–34, Jul./Aug. 2005.

[61] *End-User Multimedia QoS Categories*, ITU-T Std. Recommendation G.1010, 2001.

[62] L. Sun, “Speech quality prediction for voice over internet protocol networks,” Ph.D. dissertation, School of Computing, Communications and Electronics Faculty of Technology, University of Plymouth, Dec. 2004.

[63] S. Jelassi, H. Youssef, C. Hoene, and G. Pujolle, “Single-ended parametric voicing-aware models for live assessment of packetized voip conversations,” *Telecommunications Systems, Special Issue on the Quality of Experience issues in Multimedia Provision*, Jun. 2010.

[64] S. Jelassi and H. Youssef, “Connectivity aware instrumental approach for measuring vocal transmission quality over a wireless ad hoc network,” in *Proc. 2nd International IEEE/IFIP Conference on New Technologies, Mobility and Security (NTMS 2008)*, Tangier, Morocco, Nov.10-14 2008.

[65] X. Wang, A. Patil, and W. Wang, “Voip over wireless mesh networks: Challenges and approaches,” in *Proc. 2nd Annual ACM International Wireless Internet Conference (WICON’06)*, Boston, MA, USA, Aug.2-5 2006.

[66] K. Kim and S. Hong, “Vomesh: Voice over wireless mesh networks,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC 2006)*, Las Vegas, Nevada, USA, Apr.03-06 2006.

[67] I. F. Akyildiz, W. Y. Lee, and K. Chowdhury, “Spectrum management in cognitive radio ad hoc networks,” *IEEE Network*, vol. 23, no. 4, pp. 06–12, Jul./Aug. 2009.

[68] S. Ganguly, V. Navda, K. Kim, A. Kashyap, D. Niculescu, R. Izmailov, S. Hong, and S. Das, “Performance optimizations for deploying voip services in mesh networks,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2147–2158, Oct. 2006.

[69] P. C. Ng, S. C. Liew, and C. Lin, “Voice over wireless lan via ieee 802.16 wireless man and ieee 802.11 wireless distribution system,” in *Proc. International Conference on Wireless Networks, Communications and Mobile Computing (WIRELESSCOM’05)*, Maui, Hawaii, USA, Jun.13-16 2005.

[70] S. Madhani, S. Shah, and A. Gutierrez, “Optimized adaptive jitter buffer

design for wireless internet telephony," in *Proc. IEEE GLOBECOM'07*, Washington, DC, USA, Nov.25-30 2007.

[71] P. Vidales, "A talk on user perception of mobility in ngn," in *Proc. DTAG Workshop QoS and QoE monitoring*, Berlin, Germany, 2007.

[72] S. Mller, M. Wltermann, B. Lewcio, N. Kirschnick, and P. Vidales, "Speech quality while roaming in next generation networks," in *Proc. IEEE International Conference on Communications (ICC'09)*, Dresden, Germany, Jun.14-18 2009.

[73] H. Melvin and L. Murphy, "Exploring the extent and impact of playout adjustments within voip applications on mos scores," in *Proc. Measurement of Speech and Audio Quality in Networks (MESAQIN'05)*, Prague, Czech Republic, Jun.09-10 2005.

[74] S. Voran, "Perception of temporal discontinuity impairments in coded speech - a proposal for objective estimators and some subjective test results," in *Proc. 2nd International Conference on Measurement of Speech and Audio Quality in Networks*, Prague, Czech Republic, Jun.17-18 2003.

[75] F. Hammer, P. Reichl, and A. Raake, "Elements of interactivity in telephone conversations," in *Proc. 8th International Conference on Spoken Language Processing (ICSLP / INTERSPEECH 2004)*, Jeju Island, Korea, Oct.04-08 2004.

[76] K. Chen, C.-Y. Huang, P. Huang, and C.-L. Lei, "Qunatifying skype user satisfaction," in *Proc. ACM SIGCOMM'06*, Pisa, Italy, Sep.11-15 2006.

[77] *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 Std. Recommendation P.862, 2001.

[78] B. Sat and B. W. Wah, "Analysis and evaluation of the skype and google-talk voip systems," in *Proc. 2006 IEEE International Conference on Multimedia and Expo (ICME 2006)*, Toronto, Ontario, Canada, Jul.09-12, 2006, pp. 2153–2156.

Sofiene Jelassi was born in Bizerte, Tunisia, in 1979. He received his Bachelor of Science and Master of Science degrees in Computer Science from Faculty of Sciences of Monastir (FSM), University of Monastir, Tunisia in 2003 and 2005, respectively. He is pursuing his doctoral study between the National School of Computer Science (ENSI), University of Manouba, Tunisia and University of Pierre and Marie Curie (UPMC), Paris, France. His research interests include ad-hoc wireless networks, heterogeneous networks, multimedia content delivering, conversational application integration, seamless mobility provision, and perceptual quality assessment.

Gerardo Rubino is a Senior Researcher at INRIA, in the Center of Rennes, France. He received his Ph.D. in Computer Science from the University of Rennes 1, France, in 1985. Since then, he does research for INRIA. From 1995 to 1999, he served at the Networks and Multimedia Services Department at the ENST Bretagne Engineering School, as a responsible person for the research activities in that group. He is the head of the DIONYSOS group at INRIA, on the design and analysis of communications networks (http://www.irisa.fr/dionysos). His main interests are in the quantitative analysis of computer and communications systems, mainly using probabilistic models. He is a member of the IFIP WG 7.3 in performance evaluation. He has regular editorial and animation activities in the area, and more generally in operations research.

Hugh Melvin is based at the Discipline of Information Technology, NUI, Galway since 1997. His interests include Real-Time Systems, Internet Multimedia, Timing Systems, Wireless Technologies, Energy & Environmental Informatics. Prior to 1997, he worked for 9 years in Power Generation. Graduated with a B.E from UCD in 1988, an MBA from University of Limerick in 1997, a Research M.Sc from NUI Galway in 2000, and a PhD from University College Dublin in 2004.

Habib Youssef received a Diplôme d'Ingénieur en Informatique from the Faculté des Sciences de Tunis, University of El-Manar, Tunisia in June 1982 and a Ph.D. in computer science from the University of Minnesota, USA, in January 1990. From September 1990 to January 2001 he was a Faculty member of the computer engineering department of King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia (Assistant Professor from 1990 to 1995 and Associate Professor from September 1995 to January 2001). From February 2001 to June 2002, he was a Maitre de Conférences en informatique at the Faculté des Sciences de Monastir (FSM), University of Monastir, Tunisia. From July 2002 to August 2005, he served as the Director of the Institut Supérieur d'Informatique et Mathematiques of the University of Monastir. He is currently serving as a Professor of computer science at the Institut Supérieur d'Informatique et des Technologies de Communication, Hammam Sousse, University of Sousse, Tunisia.

Dr. Habib Youssef has over 130 publications to his credit in the form of books, book chapters, and journal and conference papers. He is the author with S. Sait of two books, (1) "VLSI Physical Design Automation: Theory and Practice," McGraw-Hill 1995, (also co-published by IEEE Press 1995), and reprinted with corrections by World Scientific in 1999, and (2) "Iterative Computer Algorithms with Applications in Engineering," IEEE CS Press 1999, and since 2003 published by John Wiley & Sons, which has also been translated into Japanese. His main research interests are computer networks, performance evaluation of computer systems, and algorithms for the design automation of electronic systems.

Guy Pujolle is currently a Professor at the Pierre et Marie Curie University (Paris 6), a member of the Institut Universitaire de France, and a member of the Scientific Advisory Board of Orange/France Telecom. Dr. Pujolle is the French representative at the Technical Committee on Networking at IFIP. He is an editor for International Journal of Network Management, WINET, Telecommunication Systems and Editor in Chief of the indexed Journal "Annals of Telecommunications." He was an editor for Computer Networks, Operations Research, Editor-In-Chief of Networking and Information Systems Journal, Ad Hoc Journal and several other journals. Guy Pujolle is a pioneer in high-speed networking having led the development of the first Gbps network to be tested in 1980. Guy Pujolle is co-founder of QoSMOS (www.qosmos.fr), Ucopia Communications (www.ucopia.com), Ginkgo-Networks (www.ginkgo-networks.com), EtherTrust (www.ethertrust.com), and Virtuor (www.VirtuOR.com).