

Jean-Yves Monfort – Klemens P. F. Adler – Hans-Wilhelm Gierlich – Joachim Pomy *

MULTIMEDIA QUALITY AS PERCEIVED BY THE USER

End-to-end voice quality as perceived by the user has always been of major concern to ETSI, the European Telecommunications Standards Institute and its Technical Committee STQ (Speech processing, Transmission and Quality aspects). Recent developments in telecommunications increasingly promote the introduction of new services, including wideband speech communication and multimedia. At the same advent the global telecommunication landscape is undergoing dramatic changes, such as the migration from traditional public network operators to internet (or NGN) service providers.

This paper provides an insight into ETSI's activities for end-to-end multimedia quality for NGN - which is a promotional factor for the market. In a multi-vendor environment with quality not being regulated, standards are the only means to achieve reasonable end-to-end multimedia quality.

There will be a huge demand for wideband speech communication and multimedia in hands-free, mobile, nomadic and video phone applications in the near future. Devices designed for such applications will have to rely on non-linear and time variant signal processing in order to be capable of providing speech quality that satisfies users' demands. Therefore, it is essential to develop state-of-the-art requirements and test methodologies and to standardize them.

1. ETSI STQ - A Center of Excellence

Within ETSI, STQ is a center of excellence for end-to-end single media and multimedia transmission performance, QoS parameters for networks and services and distributed speech recognition, and takes responsibility for related standardization of terminals and networks. STQ has a charter stretching across ALL technology platforms and thus works in close co-operation with all ETSI and 3GPP groups involved in communication aspects; this charter also reaches out to other relevant organizations, such as ITU-T, TIA and IEEE. STQ Mobile, the working group on mobile services is creating standards on QoS aspects for popular services in GSM™ and Third Generation networks including picture and video quality; new working areas include aspects of Push-to-Talk over Cellular, MTSI (Multimedia Telephony Services over IMS), mobile broadcast and the definition of a reference web page in mobile QoS.

In order to facilitate this charter the members of the STQ leading team have committed to providing leadership also in ITU-T Study Group 12 because of the close relationship of both groups.

In summary, STQ's current activities are mainly focused on all quality aspects of multimedia.

2. Motivation for Multimedia Quality

Liberalization and competition, network inter-connection, the impending change to IP technology, real-time multimedia applications and services, have brought major changes over the last few years to the way quality for telecommunications is perceived by the user. The need for a framework to test, measure performance, and achieve Media Quality and Quality of Service (QoS) that takes into account these changes has become even more necessary.

In a multi-vendor environment standards are the only means to achieve reasonable end-to-end multimedia quality. STQ represents ETSI's commitment to end-to-end multimedia quality for NGN.

It is widely recognized that end-to-end multimedia quality is a promotional factor in the market. Generally, there is a lack of requirement standards for all the elements involved in a connection supporting high media quality and even though user behavior is changing and users are apt to accept low quality in certain situations (e.g. for private chats), there is also a strong demand for high quality in other situations (e.g. for business negotiations).

* Jean-Yves Monfort, Klemens P. F. Adler, Hans-Wilhelm Gierlich, Joachim Pomy

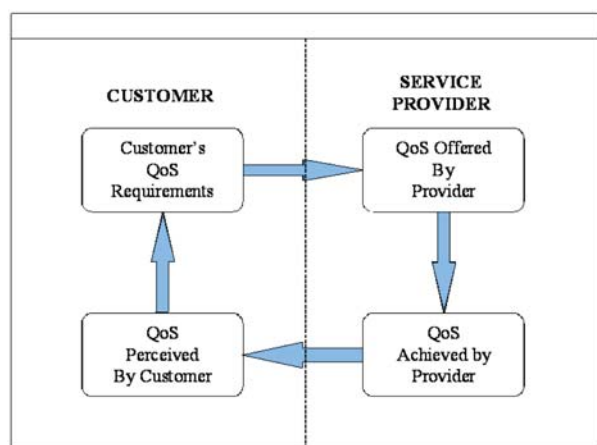
European Telecommunications Standards Institute (ETSI), Technical Committee Speech Processing, Transmission and Quality Aspects (STQ), Sophia-Antipolis Cedex, France, E-mail: jeanyves.monfort@orange-ftgroup.com
The authors jointly form the Leading Team of ETSI STQ.

3. Four Viewpoints of QoS

The QoS definition matrix of Figure 1 gives criteria for judging the quality of the communications functions that any service must support. However, even this definitional matrix can be viewed from different perspectives:

- Customer's QoS requirements (or expectations);
- Service provider's offerings of QoS (or planned/targeted QoS);
- QoS achieved or delivered;
- Customer survey ratings of QoS, i.e. the perceived QoS.

For any framework of QoS to be truly useful and practical enough to be used across the industry, it must be meaningful from these four viewpoints, which are illustrated below. While Figure 1 shows the "top down" relationship of these viewpoints, it does not indicate how, for example, QoS actually gets implemented by the service provider. This requires many detailed methods done in a more "bottom up" operation.



T1213040-02

Fig. 1 The four viewpoints of QoS (from [1])

4. Quality as perceived by the User is a Promotional Factor in the Market

Users will compare the quality of new telecommunication services offerings with the quality they have experienced in the past as well as with other telecommunications service offers. For example, for a new video conferencing service users will compare this service offering to similar service offerings by other vendors but also they will compare it to other service offerings like a pure audio conferencing.

In addition users will compare the quality of multimedia services with the quality experienced in traditional entertainment services - this may be extremely critical for new wideband and super wideband audio services.

It is also of extreme importance to understand that users have individual thresholds for quality and that they, generally speaking,

will try new services only a few times (up to 3x) and if their finding is that the perceived quality does not match their individual thresholds, they will give up and not (i.e. never) try this service again.

Users' remembrance is static in contrast to the dynamic processes of service providers. One logical consequence of this is, that users may conclude: "This website is not useable - let's try the offer of the competitor..." despite any improvement efforts of the original provider.

This may lead to the conclusion for providers of new services that a migration strategy starting from lower quality and just aiming high quality might not be sufficient to satisfy users' demands.

5. Diffusion, Transmission Quality and Expectation for an Innovation

The diffusion theory is generally accepted for describing consumer behavior on the introduction of a new service [2].

The development of expectation in a new product (an innovation) can be analyzed with the help of the diffusion theory, details of this theory can, e.g. be found in [3] and [4]. Over many studies it has been found that the number of actual users of an innovation develops in an S shaped curve, see the first diagram of Figure 2. The time it takes to diffuse a product depends on many factors, so no scaling can be given. Different people proceed through the adoption process at different points in time. According to the adoption time, users can be divided into 5 classes, see Figure 2, second diagram:

1. *Innovators*: A very small group of persons who are very quick to purchase a new product or use a service. They are very willing to accept new technologies. Innovators have been found to be people with a higher income level, higher occupational status, and they are more socially mobile than other groups. Interestingly, they are not well integrated into social groups, so they do not rely on other's opinions as to whether products suit their own purposes.
2. *Early adaptors*: A somewhat larger group following the innovators. They are still quick to purchase a product or use a service, but are much more integrated in their respective social group and believe in group norms. This is an aspect which seems to be apparent, e.g. for the early adaptors of mobile telephones.
3. The *early majority*: These people enter the market next, but they are much less willing to take risks. About one third of all adaptors belong to this group.
4. The *late majority*: This group enters the market when "newness" declines, so they are not really purchasing a new product or using a new service. They are less influenced by their corresponding social group behaviour and can be more easily influenced by advertisements.
5. The *laggards*: They enter the market when an innovation is already well accepted.

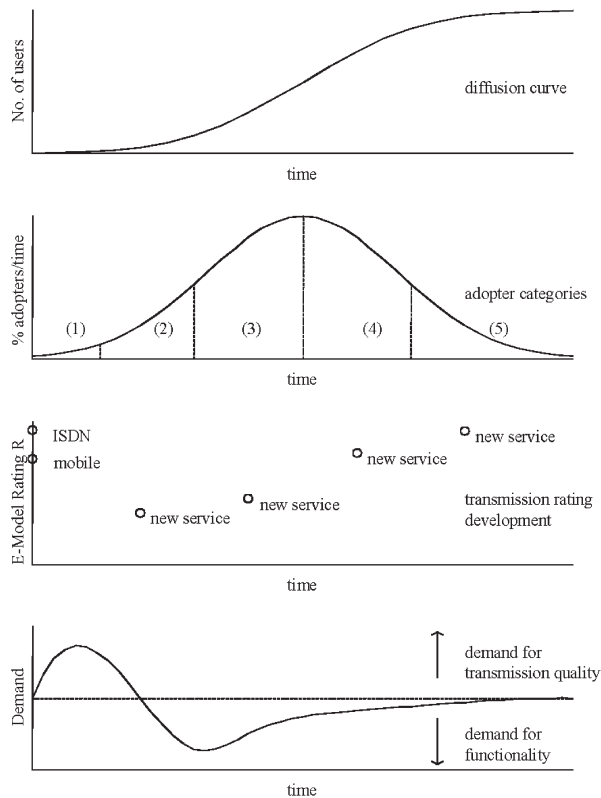


Fig. 2 Diffusion theory (from [2])

6. Where it All Begins: Real Communication Situation

When looking into users' perception of communication quality it is advised to go back to the real communication situation as it is present without technical means as depicted in Figure 3.

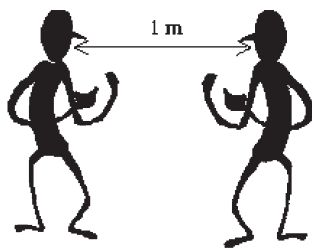


Fig. 3 Real communication situation between to human individuals

This situation has to be compared to the simulation or emulation of said situation by the means of a telecommunication system as depicted in Figure 4.

Although this classical example refers "only" to voice communication, it is easy to imagine similar comparisons for all kinds of media and multimedia communications. As one example, video on demand (VoD) will be compared by users to traditional televi-

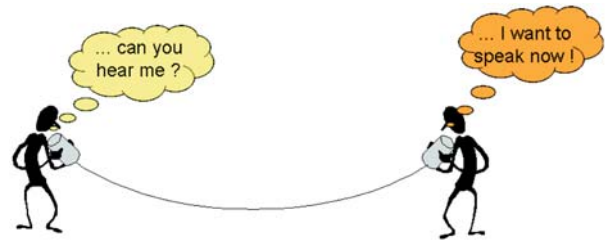


Fig. 4 Employing a telecommunication system

sion broadcast services and of course the flexibility of VoD will be honored by the user, whereas the video quality itself will be subject to a discriminating comparison with traditional TV.

7. Key Parameters affecting Multimedia Quality

In an IP environment multimedia quality will be affected mainly by the following parameters:

- Media Distortion
- End-to-End Delay
- Echo Effects
- Information Loss
- Distortion of Background Noise
- Loss of Synchronization between Media Streams

Media distortion can be recovered to a certain distinct in many cases; in detail this depends on the encoding scheme and other techniques involved. In contrast information loss which has the same root cause, can never be recovered - this is a serious problem for real-time multimedia business communication over IP. Furthermore, delay is a limited resource which is often wasted in the design of terminals or other signal processing equipment.

An increasing problem in network elements and terminals are signal processing devices which operate in tandem and thus create additional distortions (like artefacts) and delay. In order to minimize such - unwanted - quality degradations an information exchange logic and architecture concerning signal processing elements and their settings at (inter-) connection points is needed.

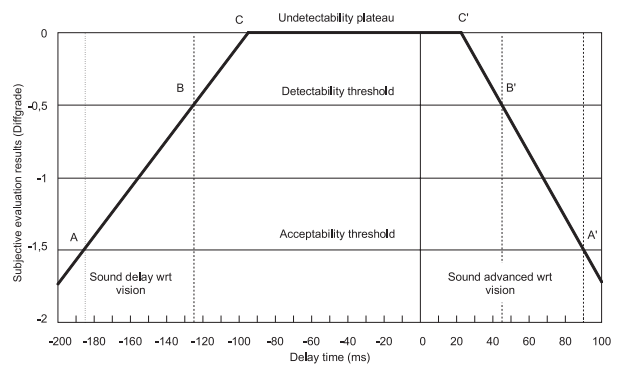


Fig. 5 Synchronization requirements (from [5]) for video and sound

A typical problem experienced in multimedia applications is the (lacking) synchronization between voice and video, see Figure 5.

Recently, German soccer fans could experience a situation where – due to a technical problem in the regular broadcasting regime – the sportscaster screamed “goal” while the viewers saw the ball somewhere in the middle field – and of course most users were unsatisfied with the (TV) service.

8. Impairments in packet networks

At each node in a packet network, packets are held in a queue awaiting transmission. Congestion causes longer queues than normal and so increases the transmission delay for the packets. Nodes also have limited queuing capacity and queues may overflow resulting in packet loss.

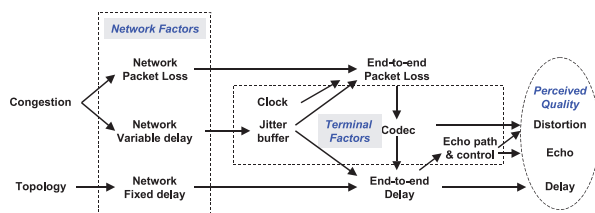


Fig. 6 Conjunction of network impairments and terminal impairments (from [6])

The terminals at the ends of speech circuits provide jitter buffering to smooth the play-out of packets. The jitter buffer is a storage of packets awaiting processing. The storage has a maximum size determined by the hardware, although the size used may be varied dynamically. The packets arrive at varying times and are extracted at regular intervals by the play-out algorithm. The effect of the jitter buffer is to convert variable delay into fixed delay. The fixed delay depends on the filling level of the jitter buffer. If the variable delay increases so that the buffer empties (called a jitter buffer under-run) then a packet is missed and cannot be played out. When variable delays reduce, the jitter buffers will fill up and if the maximum capacity is exceeded then packets will have to be discarded (called a buffer over-run). The play-out algorithm may be quite sophisticated and adjust the fill of the jitter buffer so that it is filled only to the extent necessary so that the probability of an under-run is low, so that the additional delay is minimized. In order to make these adjustments, packets may have to be skipped or duplicated introducing some additional distortion. Some highly intelligent algorithms may observe the values in the packets and make adjustments only when there are pauses in the speech flow. The different effects are shown in Figure 6. The important message here, again, is that the loss of information can be concealed but not recovered.

9. STQ Work on NGN Quality

All end-to-end Quality aspects for NGN have been moved from ETSI TISPAN to ETSI STQ, while TISPAN remains responsible for the architecture and signalling aspect to support end-to-end QoS. STQ has launched a roadmap on NGN related QoS and works – besides others – on the following topics:

- Performance Planning Guidelines for NGN Interconnect
 - NGN Interconnect Scenarios
 - End-to-End Performance Planning Guidelines for the PSTN Services in NGNs
- Audiovisual QoS for Communication over IP Networks
 - Access Technologies covered include both wired (e.g. xDSL) and wireless (e.g. UMTS, WLAN) technologies.
 - Display Size range covered is from those of small mobile terminals (e.g. 2”) up to large TV sets (e.g. 40” or more).
 - Applicable to Broadcasting and Streaming applications such as IPTV and VoD and Interactive point-to-point applications such as video telephony and videoconferencing.
- Requirements for Quality related Parameters at VoIP Interconnection Points
 - there is an urgent need in the industry for VoIP interconnect quality parameters and related requirements
- Provisional Responsiveness Thresholds for NGN Multimedia Applications from a User point of Perception
 - Possible use of the Apdex index

10. Four VoIP Terminal Standards

Four new standards specify terminal equipment requirements which enable manufacturers and service providers to deliver good quality end-to-end speech performance considering the essential requirements for the terminal equipment and their ability to handle impairments introduced by the network. These transmission requirements are drawn up from a QoS perspective as perceived by the user for

- Narrowband VoIP Terminals (handset and headset) [7]
- Narrowband VoIP Loudspeaking and Handsfree Terminals [8]
- Wideband VoIP Terminals (handset and headset) [9]
- Wideband VoIP Loudspeaking and Handsfree Terminals [10]

Besides the more traditional handset and desktop hands-free terminals also headsets, different types of hands-free communication devices ranging from handheld type to group audio terminals are considered. Furthermore setup and performance requirements for softphones are included.

A new approach was taken to make the requirements for the frequency response in sending and receiving direction more realistic; the basis is now the orthotelephonic reference response between two users in 1 meter distance under free field conditions, which is a new testing methodology developed in STQ; this methodology is compatible with the approach that has independently been taken for entertainment and multimedia devices. There are additional require-

ments for signal processing in terminals and new double talk performance requirements as well as new requirements for switching characteristics including echo cancellation tests.

11. Background Noise Transmission Quality

In modern communication the quality of the ambient background noise transmitted over the communication channel is of much more relevance than it was in the past. This is caused by three major developments:

- the increasing use of mobile and cordless devices with low D-values¹⁾, which pick up more background noise than a traditional standard terminal,
- the changing communication situations where users communicate no longer from a quiet home or small office but for example from the open-plan office, from the railway station or from the street - often with ear-deafening ambient noise,
- the increasing use and even tandeming of signal processing elements in network and terminals, such as comfort noise injection, automatic gain control or speech quality enhancement devices.

Background noise is present in most of the conversations today and it has the potential to impact the speech communication performance significantly. Therefore, testing and optimization of terminal and network equipment is necessary using realistic background noises. Furthermore reproducible conditions for the tests are required which can be guaranteed only under lab type condition.

In order to meet these challenges ETSI STQ has carried out an internal project with the support of external subjective test labs which resulted in an ETSI Guides with three parts:

- Background noise simulation technique and background noise database [11]
- Background noise transmission - Network simulation - Subjective test database and results [12]
- Background noise transmission - Objective test methods [13]

This project was funded by the European Commission since its results are essential for the high quality wideband communications of multimedia applications promoted by the EU, such as

- e-Government
- e-Health
- e-Learning
- e-Business

Based on this new standardized methods equipment manufacturers are now in the position to optimize the performance of their devices under realistic conditions based on objective test methods in the lab. Network operators and service providers are able to set minimum performance requirements based on typical use cases for their equipment which again can be defined based on the new ETSI STQ standards [11], [13].

Figure 7 shows the setup for realistic background noise simulation which can be used in labs, while Figure 8 shows a similar setup for cars. This setup can be used for all type of terminals including hands-free terminals.

The new objective test method described in [13] is based on a hearing model. It is more diagnostic than other objective speech quality tests methods since it provides in addition to the overall listening speech quality G-MOS (Global Mean Opinion Score):

- a mean opinion score purely focusing on the perceived speech quality in background noise S-MOS (Speech Mean Opinion Score) and
- a mean opinion score purely focusing on the perceived annoyance caused by background noise N-MOS (Noise Mean Opinion Score).

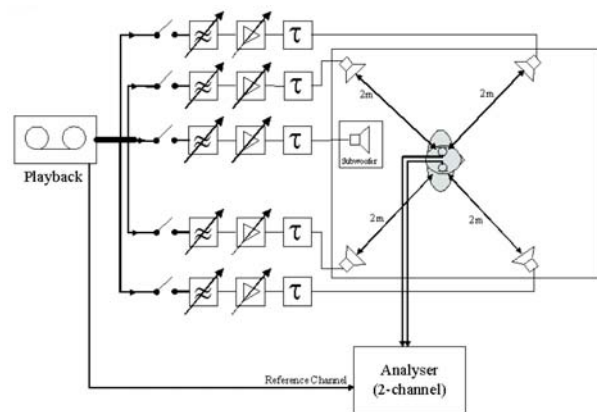


Fig. 7 Setup for realistic background noise simulation in a lab-type of environment

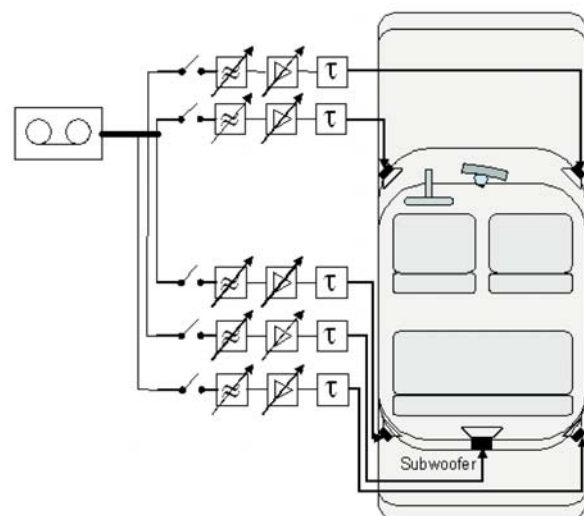


Fig. 8 Setup for realistic background noise simulation in a car

¹⁾ D-value is also referred to as Δ_{sm} , the difference in sensitivity towards diffuse noise compared with direct sound

The new objective model described in [12] is based on a huge database of a big variety of different speech samples with background noise which was investigated based on [14]. The details of the database as well as of the processing can be found in [13].

12. Conclusions

Generally, there is a lack of requirement standards for all the elements involved and interacting in a connection supporting high media quality. Even if users can deal relaxed with non-optimum quality in many cases, they have the expectation that high media quality is available to them once they demand it (and do not care so much about the actual costs). In response to these challenges

- STQ has published a report on basic issues concerning quality of speech over packet technology
- STQ is currently working on a standard for audiovisual QoS for communication over IP networks
- STQ has launched a roadmap on NGN related QoS standards to be developed in division of labor with ETSI TISPAN
- STQ has published four new terminal standards in order to enable manufacturers and service providers to enable good quality

end-to-end speech performance; the transmission requirements are drawn up from a QoS perspective as perceived by the user

- STQ has completed significant work on background noise and has developed a new methodology for testing background noise transmission.

In summary, with a charter stretching across all technology platforms STQ represents ETSI's commitment to end-to-end multimedia quality for NGN.

13 Three Excellent Reasons for Joining STQ

- Participate in the creation & improvement of standards for end-to-end media quality
- Share your own knowledge on media quality and make STQ even better
- Listen to the discussions during regular STQ sessions and special workshops & become part of the excellence. For immediate contact: info@etsi-stq.org.

References

- [1] ITU-T Rec. G.1000: Communications Quality of Service: A framework and definitions
- [2] ITU-T Rec. G.113: Transmission impairments due to speech processing
- [3] MOELLER, S.: *Assessment and Prediction of Speech Quality in Telecommunications*, Kluwer Academic Publishers, USA-Boston, 1991.
- [4] WILKIE, W. L.: *Consumer Behaviour*, John Wiley & Sons Inc., USA-New York, NY, 1994.
- [5] Draft ETSI DES/STQ-097: *Speech processing, Transmission and Quality aspects (STQ); Audiovisual QoS for communication over IP networks*
- [6] ETSI TR 102 430: *Speech Processing, Transmission and Quality Aspects (STQ); Basic Issues concerning the Quality of Speech over Packet Technology (both Internet and Next Generation Networks)*
- [7] ES 202 737: *Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user*
- [8] ES 202 738: *Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for narrowband VoIP loud-speaking and handsfree terminals from a QoS perspective as perceived by the user*
- [9] ES 202 739: *Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user*
- [10] ES 202 740: *Speech Processing, Transmission and Quality Aspects (STQ); Transmission Requirements for Wideband VoIP Loud-speaking and Handsfree Terminals from a QoS Perspective as Perceived by the User*
- [11] EG 202 396-1: *Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database*
- [12] EG 202 396-2: *Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 2: Background noise transmission - Network simulation - Subjective test database and results*
- [13] EG 202 396-3: *Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise; Part 3: Background noise transmission - Objective test methods*
- [14] ITU-T Rec. P.835: *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.*