Question 15/12 – Parametric and E-model-based planning, prediction and monitoring of conversational speech and audio-visual quality

(Continuation of Question 15/12 and parts of Question 11/12)

1   Motivation

The telecommunications industry is working to adopt more flexible infrastructure to control costs and facilitate the introduction of new services. Examples are 5G or generally next generation IP-networks which provide flexible transmission bandwidths and user interface connections, however at the expense or quality which varies with the transmission scenario and with time. A proper transmission planning, as well as flexible prediction and monitoring of Quality of Experience (QoE) are useful in managing the efficient operation and the effective services of such networks.

Regarding transmission planning of such scenarios, Study Group 12 has established the E-model, a computational model for use in transmission planning, see Recommendation G.107. This model is now frequently applied to plan traditional, narrow-band and handset-terminated networks, and to an increasing extent also for wideband, fullband telephonies and packet-based networks, using the extensions of the E-model described in Recommendations G.107.1 and G.107.2. While being popular, the E-model still shows a considerable number of limitations, namely when applying it in super-wideband and fullband networks, which non-handset terminal equipment, and with speech processing devices (such as echo cancellers, noise reduction, or alike) integrated in the network or in the terminal.

Regarding the quality prediction and monitoring of such scenarios, the industry is already benefiting from ITU-T Recommendations for objective speech quality assessment. However, most of the techniques described in these recommendations are signal based and address listening only contexts. Typical communications involve interactive, two-way, conversations. IP and mobile networks can be particularly deleterious to interactive applications, including voice conversation; for example due to increased delay, which in turn will increase the probability of double-talk and increase the perceptibility of echo. Thus, there is a need for a real-time, or near real-time, conversational speech quality assessment and monitoring.

In the end, what is needed is the integration of listening-only, talking-only and interaction quality on a common scale which could be used for planning, predicting and monitoring conversational quality in real-life networks. Such a scale would allow for an easier interpretation of the QoE provided by the different network and service scenarios, and thus make use of the flexibility offered by the respective networks in order to provide optimum services to the customer.

It is envisaged that new methods under this question would be developed collaboratively.

The following major Recommendations, in force at the time of approval of this Question, fall under its responsibility:

G.107, G.107.1, G.107.2, G.113, P.56, P.561, P.562, P.564, P.565, P.565.1, P.833, P.833.1, P.834, P.834.1

2   Question

Study items to be considered include, but are not limited to:

–   How can the E-model be used to facilitate transmission planning in wide-band, super-wideband, fullband, and mixed-band scenarios?

–   How are the relations between degradation covered by the E-model in various audio bandwidths?

– Which quality issues have to be taken into account when extending the E-model to terminal equipment other than standard handset telephones (e.g., HFTs, headsets)? Which parameters can be used to describe such terminal equipment?

– How can the perceptual effects introduced by speech-processing devices included in the network or in the terminal equipment (e.g., (acoustic) echo cancellers, level control devices, voice activity detectors, noise suppression devices) be covered by the E-model?

– Is the E-model suitable for quality monitoring? How would such a monitoring application take into account strongly time-variant channel characteristics, e.g., due to bursty frame or packet loss, or in a cellular network?

– What is the impairment effect of each new coding algorithm, so that it can be considered in the context of Recommendation G.113 and applied together with Recommendation G.107, G.107.1 or G.107.2?

– Is it possible to derive a universal quality scale which would be applicable across a range of narrowband, wideband, super-wideband and fullband scenarios, and which would integrate listening-only, talking-only and interaction aspects into one estimation of conversational call quality?

– How can non-intrusive measurements of voice quality at the IP layers be implemented and improved, for instance by taking into account signalling protocols not yet used by existing methods (e.g., SIP SDP, RTCP XR) or network technologies not covered by existing methods (mobile VoIP, WebRTC GetStats API)?

– What relationship exists between the subjective responses of users at the terminals and the objective measurements made from the point at which the non-intrusive assessment system is connected?

– What are the critical components of conversational speech and audio-visual quality? What existing models and measures addressing these components could be used as inputs and building blocks for the development of new methods?

– What subjective test methods should validation of new objective methods for the assessment of perceived conversational quality be based on?

– How can talking quality and conversational quality be measured in a non-intrusive way?

– How can existing measurement methods for voice quality be applicable for other services than telephony, in particular for video-telephony?

3   Tasks

Tasks include, but are not limited to:

– maintenance and enhancement of the E-model described in Recommendation G.107, G.107.1 and G.107.2 and input to depending Recommendations;

– frequent update of Appendices to G.113;

– maintenance of the Recommendations P.833 and P.834 and corresponding wideband and fullband Recommendations for determining equipment impairment factors;

– development of a new approach to provide a universal quality scale;

– changes and/or improvements to existing ITU-T Recommendations P.56, P.561, P.562, P.564 and P.565 to take into account new technologies;

– development of new models (both parametric and signal-based), to combine multiple objective measurements to provide an objective assessment of the perceived conversational speech and audio-visual quality;

– development of simulation-based approaches to model conversational behaviour;

– development of new models and/or relative conformance testing methodologies to assess the perceived listening and/or conversational quality of mobile IP voice and videotelephony services.

An up-to-date status of work under this Question is contained in the SG12 work programme https://www.itu.int/ITU-T/workprog/wp_search.aspx?sp=17&q=15/12.

4 Relationships

WSIS Action Lines

– C2

Sustainable Development Goals

– 9

Recommendations

– E.804, G.108, G.108.1, G.108.2, G.109, G.114, G.115, G.131, G.1050, G.1070, P.11, P.340, P.800, P.800.1, P.805, P.831, P.832, P.862, P.863

Questions

– 6/12, 7/12, 9/12, 10/12, 12/12, 13/12, 14/12, 17/12

Study Groups

– ITU-T SG16

Other bodies

– ETSI TC STQ, IETF (IPPM, XRBLOCK), TIA TR30.3