INTERNATIONAL TELECOMMUNICATION UNION

ITU-T P.880

(05/2004)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Methods for objective and subjective assessment of quality

Continuous evaluation of time varying speech quality

ITU-T Recommendation P.880

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30 P.300
Transmission standards	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	**Series**	**P.80** **P.800**
Audiovisual quality in multimedia services	Series	P.900

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation P.880

Continuous evaluation of time varying speech quality

Summary

This Recommendation describes a methodology called Continuous Evaluation of Time Varying Speech Quality (CETVSQ) that can be used for evaluating the impact of the time fluctuations of speech quality on the instantaneous perceived quality (that is perceived at any instant of a speech sequence) and on the overall perceived quality (at the end of the speech sequence). The method uses a two-part task: first, an instantaneous judgment on a continuous scale with a slider during the speech sequence, and second, an overall judgment on a standard five-category scale at the end of the speech sequence.

Source

ITU-T Recommendation P.880 was approved on 14 May 2004 by ITU-T Study Group 12 (2001-2004) under the ITU-T Recommendation A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

ITU-T Recommendation P.880

Continuous evaluation of time varying speech quality

1 Scope

This Recommendation defines a method of subjective assessment of transmitted speech quality for long speech sequences containing quality/time fluctuations. This method is based both on a continuous rating during the listening of a speech sequence and on an overall rating at the end of the speech sequence. Therefore, in addition to the measure of the overall perceived quality (as with the generally recommended methods), it provides a measure of the instantaneous perceived quality, i.e., the quality perceived at any instant of a heard transmitted speech sequence. In its current version, this method is not applicable to the selection of speech codecs. However, it is a useful tool to diagnose the effects of impairments on instantaneous and overall perceived quality, especially for discontinuous impairments with temporal fluctuations (e.g., those due to IP packet losses, handover in mobile networks, etc.). It can also aid the development and the validation of objective measurement tools that aim to predict the speech quality by detecting and analysing various types of impairments present in a speech signal; in fact, the Continuous Evaluation of Time Varying Speech Quality (CETVSQ) method can provide instantaneous judgments, as well as an overall evaluation of the subjective quality.

2 References

[1] ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

[2] ITU-R Recommendation BT-500-11 (2002), *Methodology for the subjective assessment of the quality of television pictures*.

3 Abbreviations

This Recommendation uses the following abbreviations:

ACR	Absolute Category Rating
ANOVA	ANalysis Of VAriance
CETVSQ	Continuous Evaluation of Time Varying Speech Quality
MOS	Mean Opinion Score
QoS	Quality of Service
SSCQE	Single Stimulus Continuous Quality Evaluation

4 Description of the method

4.1 Origin and motivation

The development of the Continuous Evaluation of Time Varying Speech Quality (CETVSQ) method is motivated by the fact that the Quality of Service (QoS) of new networks varies, even during a single conversation, due to specific impairments (such as, packet losses for IP, handovers in mobile networks, etc.). These impairments are characterized by transient quality artefacts (as opposed to "continuous" impairments such as signal-to-noise ratios) with a more or less high density. Short samples cannot account for this density. Moreover, because of the technical characteristics of mobile or IP networks, speech quality can vary strongly during the same communication. In order to assess speech quality, typical methods described in ITU-T

Rec. P.800 [1] use short stimuli (8 s) in subjective listening tests. These methods are well suited for time-constant speech quality. However, unless one evaluates a very large number of samples, they cannot take into consideration realistic occurrences and distribution of these impairments. In addition, they cannot take into consideration long temporal quality fluctuations for which mnesic processes (memory effects) occur that impact the overall perceived quality. The absence of a methodology for assessing time-varying speech quality motivates this Recommendation. Therefore, an assessment of the impact on perceived quality of these kinds of degradations and of their time fluctuations during a specific communication, requires speech sequences longer than those generally used in standard subjective methods. Moreover, with a method of continuous judgment, it becomes possible to study the impact of transient degradations on perceived quality at any single instant during the listening sequence, and on overall perceived quality (at the end of the sequence). This method was inspired by the method Single Stimulus Continuous Quality Evaluation (SSCQE) used in the video domain (ITU-R BT.500-11 [2]). It has also been validated for speech quality through several previous studies ([B-1], [B-2], [B-3], [B-4]).

4.2 Test preparation

4.2.1 Stimuli

The speech material should be simple, meaningful, and easy to understand. Short speech sequences should be avoided and, provisionally, speech sequence durations between 45 seconds and 3 minutes can be used. Source recordings, including recording environment and procedure, sending and recording system, talkers and speech levels could be the same as those described in B.1/P.800 [1].

4.2.2 Sources

The number and the choice of the conditions depend on the purpose of the test. The only limit is the one imposed by the test duration. If possible, it is recommended including some control conditions in the set of test conditions, i.e., conditions without any variations of physical parameters.

4.3 Listening session

4.3.1 Listeners

An adequate number (at least 24) naive listeners shall participate in the test. All the listeners shall be native speakers of the language used for the test and no listener shall have participated in a subjective experiment in the previous six months.

4.3.2 Audio presentation and testing environment

Audio presentation shall comply with the guidelines given in ITU-T Rec. P.800 [1]. These guidelines include the listening system, listening levels, and listening environment.

4.3.3 Continuous judgement recording device and set-up

An electronic slider (e.g., variable resistor) connected to a computer should be used for recording the continuous quality assessment from the subjects. This device should have the following characteristics:

– slider mechanism without any "re-set" position (i.e., no automatic return to a pre-defined position);
– linear range of travel of about 10 cm;
– fixed or desk-mounted position;
– "slider position" samples recorded twice a second (fast enough to accurately capture responses from the subjects);

– "slider position" could be coded from 0 (bottom of scale) to a minimum of 100 (top of scale), for an acceptable resolution. The initial slider position should be at the midpoint of the scale.

4.3.4 Evaluation task

For each speech sequence, the subjects' task is twofold: a continuous evaluation while listening to the sequence, and an overall evaluation at the end of the sequence. For more details and results from previous studies, see [B-1] and [B-2].

a) Continuous evaluation

Firstly, subjects are instructed to assess the speech quality of the sequence continuously by moving a slider along a continuous scale so that its position reflects their opinion on quality at that instant; the subjects can position the slider anywhere on the scale. Five labels are shown along the scale, i.e., Excellent, Good, Fair, Poor and Bad to help the subject associate the slider position with suitable ranges of speech quality.

Continuous-quality scale

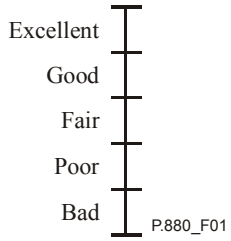

Figure 1/P.880 – Continuous scale used for the instantaneous judgment

b) Overall evaluation

Secondly, at the end of each sequence, subjects are asked to rate its overall quality on the following 5-category listening-quality scale (the same MOS scale used in the ACR).

Overall-quality scale (ACR)

Quality of the speech	**Associated score**
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

4.3.5 Test procedure

Prior to the assessment of the test speech sequences, subjects undergo training by listening to a few selected sequences. Training sequences should cover different quality levels and different quality fluctuations representative of the range of temporal fluctuations and quality levels that the subjects will encounter during the actual test.

Generally the test consists of a number of sessions, separated with breaks. The entire set of stimuli (sequences) is presented in a different random order to the different groups of subjects.

4.3.6 Instructions to subject

An example of typical instructions is given in Table 1 (for speech sequences of duration T seconds). The written instructions must be given (verbally as well, if necessary) prior to the beginning of the experiment.

Table 1/P.880 – Example of instructions that would be given to the listeners

In this test, you will be listening to T-s speech sequences via the telephone handset. The speech quality of each sequence can vary in time in different ways. For each sequence, your tasks are: 1) To give your opinion on speech quality, during the entire sequence, i.e., at any instant, by moving the slider on the table in front of you. The rating scale is continuous so that you can place the slider anywhere, with its position reflecting your opinion of the speech quality. Labels are shown along the scale (Excellent, Good, Fair, Poor and Bad) and are provided to help you for positioning the slider. For example, if you think that the quality corresponds exactly to a position between Fair and Good, you will place the slider in the middle of the two corresponding labels. However, you do not have to position the slider either directly in front of a label or exactly half-way between two labels unless one of those positions accurately represents your opinion. Do not forget to move the slider as the quality varies. 2) At the end of each sequence, you are asked to give an overall quality score that should reflect your opinion on the speech quality of the entire sequence. You give this score by pressing the appropriate button to indicate your opinion on the following scale: **Overall opinion on the speech quality of the sequence you have just heard**: Excellent Good Fair Poor Bad You will have five seconds to record your answer by pushing the button corresponding to your choice. There will be a short pause before the presentation of next sequence. We will begin with a short practice session to familiarize you with the test procedure. The actual tests will take place during sessions of 10 to 15 minutes. Thank you for your help.

4.4 Statistical analysis

For each subject, if T-s corresponds to the duration of each sequence (in seconds), a data file of 2 × T-s values is recorded (i.e., one instantaneous score every 500 ms during T seconds), plus one scalar value (i.e., the overall quality judgment). The 2 × T-s instantaneous values (from $t = 0$ until $t = 2T - 1$) are subsequently linearly transformed into values from 1 to 5, using the relation $S(t) = 1 + 4$ (slider position/maximum), where $S(t)$ is the instantaneous opinion score.

For each sequence, a mean instantaneous judgment (and its standard deviation) is obtained by averaging individual instantaneous judgments over the subjects, at each instant t (i.e., every 500 ms). For each sequence, a mean overall judgment MOS (and its standard deviation) is obtained by averaging individual overall judgments over the subjects on the ACR scale.

Substantial deviations between the continuous score and the overall score are likely indications that transient impairments were experienced. This will depend on the recency, number of occurrences and duration of the impairments relative to the overall score judgement. Further study is required before this information can be used in transmission planning.

Statistical analysis (e.g., ANOVA) can be performed to identify significant effects present in the different experimental conditions.

Remark: The responses from some subjects can be discarded if those responses exhibit high variations and appear for more than 10% of the time out of two inter-subject standard deviations (calculated over all subjects).

BIBLIOGRAPHY

[B-1] Contribution ITU-T [COM 12-94] (1999), *Continuous assessment of time-varying subjective vocal quality and its relationship with overall subjective quality*, (Source FT R&D).

[B-2] GROS (L.), CHATEAU (N.), Instantaneous and Overall Judgments for Time-Varying Speech Quality: Assessments and Relationships, *Acta Acustica*, 87, 367-377, 2001.

[B-3] Contribution ITU-T [COM 12-D139] (2000), *Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences: influence of recency effect*, (Source FT R&D).

[B-4] Contribution ITU-T [COM 12-22] (2000), *Assessment of time-varying speech quality: comparison between listening situation and conversational situation*, (Source FT R&D).

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure, Internet protocol aspects and Next Generation Networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2004