ETSI TS 103 801 V1.1.1 (2020-11)

Speech and multimedia Transmission Quality (STQ); Subjective test methodologies for the evaluation of echo control systems

Reference

DTS/STQ-285

Keywords

conversation, double talk, echo, impairment, listening quality, test

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

In speech communication devices of all kinds, echo artefacts and double talk impairments can occur. These might dramatically degrade a conversation between users, i.e. the quality of experience in general. With an increasing usage of hands-free terminals (e.g. motor vehicle, handheld or desktop devices) and new types of devices supporting voice services (e.g. smart home devices or wearables), the cancellation of echo and providing duplex communication at the same time is still a challenging task for signal processing components.

The objective assessment of degradations caused by echo and/or poor double talk performance is already covered in several specifications, but mainly based on simple analyses in level or spectrum. The impact on the conversation as perceived by the user is typically rarely investigated.

The auditory evaluation of a conversation between two human test subjects in a laboratory may be quite cumbersome. Even though some listening test specifications already exist in several standardization bodies for these scenarios, the reproducibility of results may vary a lot due to several degrees of freedom, e.g. a randomly degraded communication channel or usage of free speech.

The present document provides a subjective test framework for the evaluation of echo artefacts and double talk impairments, based on the Third-Party Listening Test (TPLT) approach. On one hand, a conversation is simulated as close as possible to human perception, in particular including the acoustics of involved terminals as well as self-hearing and self-masking in talking phases. On the other hand, the proposed test methodology utilizes pre-recorded signals, designed with respect to best-possible reproducibility in listening labs. This approach is well known from classical subjective evaluations of speech, audio and/or video. This leads to a decreased naturalness and spontaneity compared to a real conversation between subjects. However, the compromise between these two opposite approaches provides a wider range of use cases. In addition, the signals used for subjective testing may be re-used for predictive models.

1 Scope

The present document provides a framework for auditory testing of echo artefacts and double talk impairments that may occur in telecommunication devices of all kind.

The present document assesses degradations in end-to-end scenarios as perceived by the listener at the reference-side. Only degradations caused by the terminal located at the device-side are taken into account by the framework. Since the network delay between reference-side and device-side (and vice-versa) also has an impact on the DUT's signal processing and/or the listener's quality of experience, this parameter is included in the present document as well - any other degradations (e.g. packet-loss in one of the two directions) are out of scope.

Only DCR scales are supported in the auditory test, in particular for echo artefacts and double talk disturbances, which have the most impact on conversations (more may be added in the future). ACR scales e.g. speech distortion or overall quality are not considered for auditory testing.

Any instrumental model predicting results according to the introduced listening test design is out of scope.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents, which are not found to be publicly available in the expected location, might be found at https://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents are necessary for the application of the present document.

[1] Recommendation ITU-T P.10/G.100: "Vocabulary for performance, quality of service and quality of experience".

[2] Recommendation ITU-T P.800: "Methods for subjective determination of transmission quality".

[3] Recommendation ITU-T P.831: "Subjective performance evaluation of network echo cancellers".

[4] ITU-T Handbooks: "Handbook of subjective testing practical procedures".

[5] Recommendation ITU-T P.805: "Subjective evaluation of conversational quality".

[6] Recommendation ITU-T P.700: "Calculation of loudness for speech communication".

[7] ETSI TS 103 737: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

[8] ETSI TS 103 738: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

[9] ETSI TS 103 739: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

[10] ETSI TS 103 740: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

[11] ETSI TS 102 924: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for Super-Wideband / Fullband handset and headset terminals from a QoS perspective as perceived by the user".

[12] ETSI TS 102 925: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for Super-Wideband / Fullband handsfree and conferencing terminals from a QoS perspective as perceived by the user".

[13] Recommendation ITU-T P.57: "Artificial ears".

[14] Recommendation ITU-T P.58: "Head and torso simulator for telephonometry".

[15] Recommendation ITU-T P.64: "Determination of sensitivity/frequency characteristics of local telephone systems".

[16] ETSI TS 126 132 "Universal Mobile Telecommunications System (UMTS); LTE; Speech and video telephony terminal acoustic test specification (3GPP TS 26.132)".

[17] Recommendation ITU-T P.501: "Test signals for use in telephony and other speech-based applications".

[18] Recommendation ITU-R BS.708: "Determination of the electro-acoustical properties of studio monitor headphones".

[19] IEC 60268-7:2010: "Sound system equipment - Part 7: Headphones and earphones".

[20] ETSI TS 103 281: "Speech and multimedia Transmission Quality (STQ); Speech quality in the presence of background noise: Objective test methods for super-wideband and fullband terminals".

[21] Recommendation ITU-T P.56: "Objective measurement of active speech level".

[22] ETSI ES 202 737: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

[23] ETSI ES 202 738: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[24] ETSI ES 202 739: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

[25] ETSI ES 202 740: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[26] Recommendation ITU-T G.191: "Software tools for speech and audio coding standardization".

[27] Recommendation ITU-T P.79: "Calculation of loudness ratings for telephone sets".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] F. Kettler, H.-W. Gierlich, E. Diedrich and J. Berger: "Echobeurteilung beim Abhören von Kunstkopfaufnahmen im Vergleich zum aktiven Sprechen", DAGA Conference, Hamburg, 2001.

[i.2] Recommendation ITU-T P.835: "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm".

[i.3] Recommendation ITU-T P.76: "Determination of loudness ratings; fundamental principles".

[i.4] ETSI TR 126 931: "Universal Mobile Telecommunications System (UMTS); LTE; Evaluation of Additional Acoustic Tests for Speech Telephony (3GPP TR 26.931)".

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in Recommendation ITU-T P.10/G.100 [1] and the following apply:

attribute: description of a certain quality dimension of a stimulus, which is auditorily assessed by subjects in a listening test (e.g. annoyance of echo)

NOTE: Multiple attributes may be assessed for a single stimulus within one trial.

category: magnitude, which quantifies the degree of quality or degradation within an attribute

NOTE: The meaning of a certain category may be expressed by labels/descriptions, numbers or graphical alignment in the voting console to the test subject.

device-side: end-point of a telecommunication connection, which is dedicated to and operated by a device under test

NOTE: For the signal-based TPLT, a HATS is used here in order to cause double talk.

double talk: phase within a conversation (or a speech-based test signal), where the user B/ reference side as well as the user A/ DUT side are talking

double talk impairment: audible degradation in terms of quality and/or intelligibility, which is inserted by the device-side and is perceived by the listener at the reference-side

NOTE: Technically, it is typically caused by the simultaneous talker activity of both sides.

double talk source signal: signal originated from device-side and transmitted to reference-side

echo artefact: artefact generated by the signal processing in sending direction of the device-side (e.g. due to linear/non-linear coupling of signal components from receiving to sending direction of the device under test)

NOTE: It is triggered in talking phases of the reference-side.

reference-side: end-point of a telecommunication connection, which is operated by a reference device or gateway in order to capture stimuli for a TPLT

NOTE: This may be realized either electrically or acoustically with a HATS.

scale: list of categories, sorted by the degree of quality or degradation for a given attribute

signal under test: signal transmitted from device-side to reference-side

NOTE: May contain echo artefacts and/or double talk impairments caused by signal processing of DUT.

Single Talk (ST): phase within a conversation (or a speech-based test signal), where only one side/end is talking (either user B/ reference side or user A/ DUT side) is talking

source signal: signal originated from reference-side and transmitted to device-side

NOTE: May also be inserted electrically at POI to the DUT.

3.2 Symbols

For the purposes of the present document, the symbols given in Recommendation ITU-T P.10/G.100 [1] and the following apply:

a_{DT}	Attenuation (in dB) during double talk segments
g_{EL}	Factor (in dB) to obtain a certain echo loss
δ(k)	Dirac impulse (linear transmission)
ΔT	Duration of delay introduced in the echo path
dB	deciBel
dB_{SPL}	Sound Pressure Level in dB, referenced to 20 µPa
dB_{Pa}	Sound Pressure Level in dB, referenced to 1 Pa
dB_{V}	Voltage in dB, referenced to 1 Volt
$dB_{Pa/V}$	Sensitivity in receiving direction (Pascal per Volt), expressed in dB
$dB_{V/Pa}$	Sensitivity in sending direction (Volt per Pascal), expressed in dB
h(k)	Impulse response of echo path
Pa	Pascal (pressure)
T_C	Duration of concurrent talk (uplink and downlink active)
T_D	Duration of activity in downlink path
T_L	Duration of long interrupts
T_P	Duration of trailing and leading pause
T_S	Duration of short interrupts
T_U	Duration of activity in uplink path
x(k)	downlink signal sent to Device-side
x_{ST}(k)	Sidetone signal based on x(k)
y(k)	uplink signal sent by Device-side

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in Recommendation ITU-T P.10/G.100 [1] and the following apply:

5G NR	5G New Radio
ACR	Absolute Category Rating
AEC	Acoustic Echo Control
ASL	Active Speech Level
CT	Conversational Test
DCR	Degradation Category Rating
DT	Double Talk
DUT	Device Under Test
ES	Echo Suppression
FB	FullBand (20 Hz to 20 kHz)
FIR	Finite Impulse Response
GSM	Global System for Mobile Communications
INF	Infinity
IP	Internet Protocol
LTE	Long Term Evolution
NB	NarrowBand (300 Hz to 3 400 kHz)
NR	Noise Rating
NS	Noise Suppression
POI	Point of Interconnection
RCV	Receiving Direction
SLR	Sending Loudness Rating
SND	Sending Direction
SPL	Sound Pressure Level
ST	Single Talk
SWB	Super-wideband (50 Hz to 14 kHz)
TALT	Talking And Listening Test
TPLT	Third-Party Listening Test
UMTS	Universal Mobile Telecommunications System
VoIP	Voice-over-IP

WB WideBand (100 Hz to 7 kHz)

4 Fundamentals of acoustic echo control characteristics

4.1 Overview

Figure 1 depicts the simplified technical principles and components of a bidirectional end-to-end speech communication between user A (left) and user B (right). On each side, a terminal with electric and acoustic send and receive path is used. Both paths include several signal processing blocks like AEC, ES, NR, AGC and codec. The acoustic paths may range from handsets close to the ear up to recent hands-free application. The devices transmit voice signals over arbitrary and cascaded networks (e.g. VoIP access, mobile network or even satellite link).

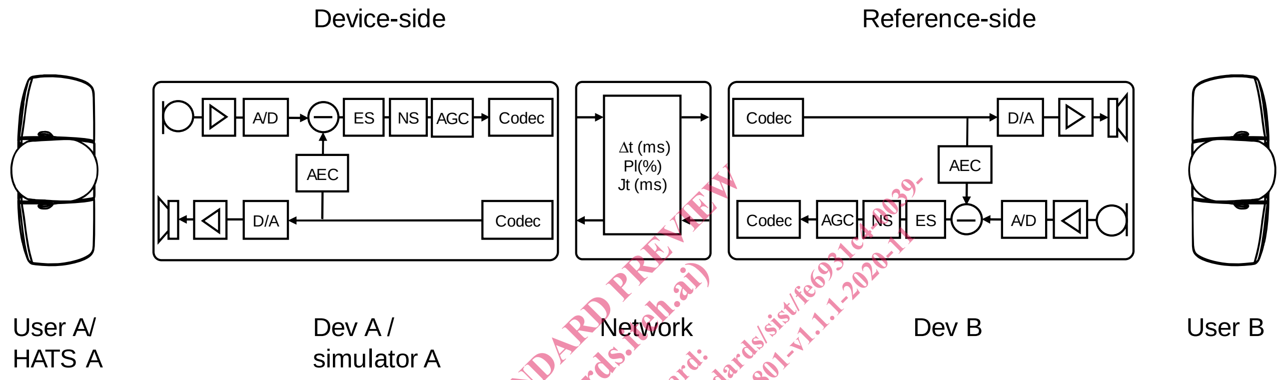

Figure 1: Technical scheme of conversation in telecommunication

NOTE: In the present document, the specific type of network is in general of minor relevance, since the degree of degradations mostly depends on the delay. However, network-specific features (e.g. coding and decoding of speech signal) should be regarded whenever possible.

4.2 Formation of Echo Artefacts

In the following, echo artefacts are described from the perspective of user B (reference side), as illustrated in Figure 3. User B starts talking and the reference device transmits the signal in sending direction via the network where delay, jitter and packet loss are possibly inserted. The signal is then played back at the device side (e.g. by loudspeaker or handset) and coupled back into the DUT's microphone. Here typically signal-processing components like an (acoustical) echo canceller and/or suppressor try to remove the echo signal. Any remaining signal is called *residual echo*, which may be even further degraded by the following signal processing units (NS, AGC, etc.).

The residual echo is transmitted back to the reference device via the network and played back to user B. In general, the resulting residual echo to be perceived by user B may be a delayed, attenuated and (linearly and/or non-linearly) distorted version of the source signal transmitted by user B. Since the roundtrip delay of the whole transmission is typically in the range of (at least) a few hundred milliseconds, user B may perceive already an echo signal while he/she is still talking. In this case, the echo signal may be partially masked by the sidetone of his/her own voice.