ETSI TS 102 924 V1.2.1 (2018-03)

TECHNICAL SPECIFICATION

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

Reference
RTS/STQ-208-1

Keywords
QoS, terminal

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 only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

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

Speech terminals are currently implementing narrowband and wideband bandwidth. Terminal equipment may offer wider bandwidth, due to features already available in these terminals. Such equipment may implement conversational features that may benefit of the electroacoustic equipment already available in the terminal and may provide wider quality for the end users.

The present document is intended to provide initial requirements and test methods for such type of equipment.

1 Scope

The present document provides speech & audio transmission performance requirements and measurement methods for handset and headset functions of super-wideband/fullband terminals. The present document provides requirements in order to optimize the end to end quality perceived by users.

Users become more sensitive to voice and music quality (for music used in conversational services) when using ICT/terminal equipment and so are more demanding for further enhancement especially further extension of the audio coded bandwidth.

For instance, this is the case for high quality conferencing services with music on hold, better background environment rendering and longer duration than normal point to point calls.

Standardized super-wideband and fullband coders are now available, some being also compatible with wideband coders.

The present document will consider only conversational services (that may be mixed with other services) and does not cover the streaming-only services.

Such applications include:

- Speech and audio communication including conferencing.
- Bandwidth extension which may allow usage for some mixed content.
- Super-wideband enhancement coupled with stereo/dichotic.

The send path it can be characterized in two ways:

- The signal picked up by microphone may combine speech, music and every type of environmental signal.
- Direct insertion of any type of signal.

For receive path, signal may be combine two types:

- Communication signals such as described for send path.
- Signal coming from distributed applications (e.g. advertisement, music on hold, etc.).

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.501: "Test signals for use in telephonometry".

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

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

[4] Recommendation ITU-T P.581: "Use of head and torso simulator for hands-free and handset terminal testing".

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

[6] Recommendation G.711.1 (2008) Amendment 4 (11/10): "Wideband embedded extension for G.711 pulse code modulation".

[7] Recommendation ITU-T G.722.1 (annex C): "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[8] Recommendation G.729.1 (05/06): "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729".

[9] Recommendation ITU-T G.718 (06/08)": "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s".

[10] Recommendation ITU-T G.719: "Low-complexity, full-band audio coding for high-quality, conversational applications".

[11] ETSI TS 103 224: "Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database".

[12] 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".

[13] 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".

[14] Recommendation ITU-T P.863: "Perceptual objective listening quality assessment".

[15] Recommendation ITU-T P.380: "Electro-acoustic measurements on headsets".

[16] IEC 61260-1: "Electroacoustics - Octave-band and fractional-octave-band filters - Part 1: Specifications".

[17] Void.

[18] Void.

[19] Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".

[20] Void.

[21] Recommendation ITU-T G.711.1 (annex F): "Wideband embedded extension for G.711 pulse code modulation".

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

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

[24] ISO 3745: "Acoustics -- Determination of sound power levels and sound energy levels of noise sources using sound pressure -- Precision methods for anechoic rooms and hemi-anechoic rooms".

[25] ETSI TR 126 952: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); Performance characterization (3GPP TR 26.952 version 12.2.0 Release 12)".

[26] ETSI TS 126 441: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); General overview (3GPP TS 26.441)".

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

[28] 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".

[29] Recommendation ITU-T G.122: "Influence of national systems on stability and talker echo in international connections".

[30] Recommendation ITU-T P.340: "Transmission characteristics and speech quality parameters of hands-free terminals".

[31] Recommendation ITU-T P.502: "Objective test methods for speech communication systems using complex test signals".

[32] Recommendation ITU-T P.863.1: "Application Guide for Recommendation ITU-T P.863".

[33] Recommendation ITU-T P.1010: "Fundamental voice transmission objectives for VoIP terminals and gateways".

[34] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".

[35] 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".

[36] IETF RFC 6716: "Definition of the Opus Audio Codec".

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] ISO 532: "Acoustics -- Method for calculating loudness level".

[i.2] NIST Net™.

NOTE: Available at https://www-x.antd.nist.gov/itg/nistnet/.

[i.3] Netem™.

NOTE: Available at http://www.linuxfoundation.org/en/Net:Netem.

[i.4] Trace Control for Netem (TCN) (2006): "Trace Control for Netem, Semester Thesis SA-2006-15", ETH Zürich, A. Keller.

[i.5] ETSI EG 202 425: "Speech Processing, Transmission and Quality Aspects (STQ); Definition and implementation of VoIP reference point".

[i.6] STQ(15)48-0309: "Objective Codec Evaluation of EVS. HEAD acoustics GmbH".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

binaural listening: both ears are involved for the perception of sound

dichotic: relating to or involving the presentation of a stimulus to one ear that differs in some respect (as pitch, loudness, frequency, or energy) from a stimulus presented to the other ear

diotic: pertaining to or affecting both ears (same signal in both ears)

dual channel mode: audio mode, in which two audio channels with independent programme contents (e.g. bilingual) are encoded within one audio bit stream

fullband bandwidth: transmission of speech with a nominal bandwidth of 20 Hz - 20 kHz

stereo mode: audio mode in which two channels forming a stereo pair (left and right) are encoded within one bit stream and for which the coding process is the same as for the Dual channel mode

super-wideband: transmission with supre-wideband bandwith which may cover at least mono capabilities. Stereo capabilities may be possible

super-wideband bandwidth: transmission of speech with a nominal pass-band wider than 100 Hz to 7 000 Hz, usually understood to be 50 Hz - 14 000 Hz (definition from Recommendation ITU-T P.10 /G.100 [2])

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACR	Absolute Category Rating
DRP	ear Drum Reference Point
ERP	Ear reference Point
EVS	Enhanced Voice Services
FB	FullBand
GAT	Group Audio Terminal
G-MOS-LQO$_F$	Overall Quality Mean Opinion Score, Listening Quality Objective, fullband
HATS	Head and Torso Simulator
MCU	Multiplexing Control Unit
MRP	Mouth Reference Point
MS	Mid-sized Stereo
N-MOS-LQO$_F$	Noise Quality Mean Opinion Score, Listening Quality Objective, fullband
POI	Point Of Interconnection
SLR	Send Loudness Rating
S-MOS-LQO$_F$	Speech Quality Mean Opinion Score, Listening Quality Objective, fullband
SWB	Super-WideBand
TCL	Terminal Echo Loss

4 Applications and coder considerations

4.1 Applications

The following applications are within the scope of the present document:

- Speech and audio communication including conferencing using high quality hands free systems, for which super-wideband/fullband coding can better reproduce the audio environment and provide improved quality and audio immersion. These applications cover also GATs (Group Audio Terminals) and teleconference systems such as "Telepresence".
- Bandwidth extension which may allow usage for some mixed content applications where wider bandwidth could bring a significant added value for the customer (support of 14 kHz and 20 kHz bandwidth and stereo/multichannel capability).
- Super-wideband enhancement coupled with stereo/multichannel to maximize the quality enhancement for the customer when the terminal device can support this capability.

The send path can be characterized in two ways:

- The signal picked up by microphone(s) may combine speech, music and every type of environmental signal.

NOTE: For some applications (e.g. journalist reporting) the user should have the possibility to cancel the noise environment or to transmit it without degradation.

- Direct insertion of any type of signal.

For receive path, signal may combine the two following types:

- Communication signal such as described for send path.
- Signal coming from distributed applications (e.g. advertisement, music on hold, etc.).

4.2 Coder considerations

4.2.0 Premise

As indicated in the scope only coders supporting conversational SWB and FB services are applicable to the present document.

4.2.1 Super-wideband (SWB)

Table 0: Use cases for coders

Coder Reference	Speech	Other signals	Stereo	Remark
VoLTE (IMS) ETSI TS 126 441 [26]	X	X Music	X	
Recommendation ITU-T G.722.1 [7] annex C	X	X Music		For low frame loss
Recommendation ITU-T G.729.1 [8] annex E (extension SWB)	X	X background noise (X) Music		
Recommendation ITU-T G.718 [9] annex B	X	X Music		
Recommendation ITU-T G.711.1 annexes D [6] and F [21]	X	X	X (annex F)	
Recommendation ITU-T G.722 [19] annexes B and D	X	X	X (annex D)	
OPUS [36]	X	X	X	
NOTE: G 722.1 [7] is intended to be used for hand-free application. It is referenced here considering that a terminal using this coder may implement a handset or headset function.				

When X is in brackets, it means that the coder is not optimized for this application.

The following coders are recommended for Super-wideband:

- Recommendation ITU-T G.722.1 [7] Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss. Annex C 14 kHz mode at 24, 32 and 48 kbit/s.
  The algorithm is recommended for use in hands-free applications such as conferencing where there is a low probability of frame loss. It may be used with speech or music inputs. The bit rate may be changed at any 20 ms frame boundary. New annex C contains the description of a low-complexity extension mode to G.722.1, which doubles the algorithm to permit 14-kHz audio bandwidth using a 32-kHz audio sample rate, at 24, 32, and 48 kbit/s.
  Annex C of [7]: this annex provides a description of the 14-kHz mode at 24, 32 and 48 kbit/s for this Recommendation.

- Recommendation ITU-T G.729.1 [8] annex E (extension SWB for G.729.1 [8]).
  This annex provides the high-level description of the higher bit-rate extension of G.729 designed to accommodate a wide range of input signals, such as speech, with background noise and even music.

- Recommendation ITU-T G.718 [9] annex B Super-wideband scalable (extension for Recommendation ITU-T G.718 [9]). This annex describes a scalable super-wideband (SWB, 50 to 14 000 Hz) speech and audio coding algorithm operating from 36 to 48 kbit/s and interoperable with Recommendation ITU-T G.718 [9].

- Recommendation ITU-T G.711.1 [6] annex D defines the super-wideband extension
Annex F defines the Stereo embedded extension for Recommendation ITU-T G.711.1 [6].
"Annex F is intended as a stereo extension to the G.711.1 [6] wideband coding algorithm and its super-wideband annex D. Compared to discrete two-channel (dual-mono) audio transmission, this stereo extension G.711.1 [6] annex F saves valuable bandwidth for stereo transmission. It is specified to offer the stereo capability while providing backward compatibility with the monaural core in an embedded scalable way. The annex provides very good quality for stereo speech contents (clean speech and noisy speech with various stereo sound pickup systems: binaural, MS, etc.), and for most of the conditions it provides significantly higher quality than low bitrate dual-mono. For some music contents, e.g. highly reverberated and/or with diffuse sound, the algorithm may have some performance limitations and may not perform as good as dual-mono codecs, however it achieves the quality of state-of-the-art parametric stereo codecs".

- Recommendation ITU-T G.722 [19] annex B defines the super-wideband extension
and annex D defines the Stereo embedded extension for Recommendation ITU-T G.722 [19].
"Annex B describes a scalable super-wideband (SWB, 50 to 14 000 Hz) speech and audio coding algorithm operating at 64, 80 and 96 kbit/s. The Recommendation ITU-T G.722 [19] super-wideband extension codec is interoperable with Recommendation ITU-T G.722 [19]. The output of the Recommendation ITU-T G.722 [19] SWB coder has a bandwidth of 50 to 14 000 Hz".
"Annex D describes a stereo extension of the wideband codec G.722 and its super-wideband extension, G.722 annex B. It is optimized for the transmission of stereo signals with limited additional bitrate, while keeping full compatibility with both codecs. Annex D operates from 64 to 128 kbit/s with four super-wideband stereo bitrates at 80, 96, 112 and 128 kbit/s and two wideband stereo bitrates at 64 and 80 kbit/s".

- 3GPP VoLTE (IMS) ETSI TS 126 441 [26]. The Enhanced Voice Services coder consists of the multi-rate audio coder optimized for operation with voice and music/mixed content signals, a source controlled rate scheme including a voice/sound activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

  Coder EVS (Enhanced Voice Services) is defined in ETSI TS 126 441 [26] and ETSI TR 126 952 [25]. The tests conducted on codec implementations, e.g. [i.6] show that the requirements and test methods for SWB terminals as defined in the present document apply.

  EVS is designed for packet-switched and circuit-switched networks/Mobile VoIP and VoLTE is a key target application.
  The key features of EVS are Super-wideband speech (32 kHz sampling) with improved speech quality and improved music performance.

4.2.2 Fullband (FB)

The following codecs are recommended for fullband:

- Recommendation ITU-T G.719 [10] Low-complexity, fullband audio coding for high-quality, conversational applications.
"Recommendation ITU-T G.719 [10] describes the G.719 [10] coding algorithm for low-complexity fullband conversational speech and audio, operating from 32 kbit/s up to 128 kbit/s".

  The encoder input and decoder output are sampled at 48 kHz. The codec enables full bandwidth, from 20 Hz to 20 kHz, encoding of speech, music and general audio content. The codec operates on 20-ms frames and has an algorithmic delay of 40 ms.

NOTE: Amendment 1 adds new annex A that specifies the use of the ISO base media file format as container for the G.719 [10] bitstream addresses non-conversational use cases of the codec (e.g. call waiting music playback and recording of teleconferencing sessions, voice mail messages, online "jam"-sessions).

- 3GPP VoLTE (IMS) ETSI TS 126 441 [26]. The Enhanced Voice Services coder consists of the multi-rate audio coder optimized for operation with voice and music/mixed content signals, a source controlled rate scheme including a voice/sound activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

Coder EVS (Enhanced Voice Services) is defined in ETSI TS 126 441 [26] and ETSI TR 126 952 [25]. The tests conducted on codec implementations, e.g. [i.6] show that the requirements and test methods for FB terminals as defined in the present document apply.

EVS is designed for packet-switched and circuit-switched networks/Mobile VoIP and VoLTE is a key target application. The key features of EVS are Fullband speech with improved speech quality and improved music performance.

5 Test considerations and test equipment

5.0 Introduction

The terminals within the scope of the present document are not only dedicated to speech communication but are also mixing speech and audio contents and may implement stereo and multichannel transmissions. As a consequence there is a need to define new parameters, such as:

- **Loudness:** Loudness Rating is determined only for speech or speech-like signals. Loudness may be calculated over any types of signals (audio sequences, speech sequences and mix of these sequences). Moreover it is not intended to define Loudness Rating algorithms for Super-wideband and fullband speech. To be consistent with transmission planning, the loudness rating shall be determined using wideband calculation and loudness shall be measurement for all the bandwidths. Clause 5.4.1.2 details the measurement principles.
- **Binaural listening:** The most of the test assessment methods and requirements for speech terminals are based on monaural listening, Even if some of them (e.g. for Handsfree Loudness rating) are intended to take into account binaural listening, the basic methods and requirements are only taking into account correction factors. The plan is to adapt test methods to effective binaural listening.

As a consequence, the present document takes into account test arrangements that are defined for speech terminals or for audio equipment.

Recommendation ITU-T P.58 [3] give information about use of HATS only from 100 Hz to 10 kHz, but new designs offer wider bandwidths.

For send the HATS can be used between 50 Hz and 16 kHz. Until the development of new systems with larger bandwidth, send measurement will be limited to those frequencies.

NOTE 1: With some measurement equipment the use of such of bandwidth is not possible and should be limited to 100 Hz to 14 kHz.

For receive, a correction factor (given, in annex B) allows measurement at DRP until 16 kHz.

NOTE 2: It is not the intention of the present document to define new requirements to adapt HATS for super-wideband and fullband. However when terminals implement Super-wideband or Fullband within terminals support also WideBand and/or NarrowBand speech, it is intended to use as far as possible test methods defined for wideband terminals and consequently to use HATS for parameters measured in wideband bandwidth.

5.1 IP half channel measurement adaptor

The IP half channel measurement adaptor is described in ETSI EG 202 425 [i.5].

5.2 Environmental conditions for tests

The following conditions shall apply for the testing environment:

a) ambient temperature: 15 °C to 35 °C (inclusive);

b) relative humidity: 5 % to 85 %;