ETSI TR 103 138 V1.4.1 (2016-09)

TECHNICAL REPORT

**Speech and multimedia Transmission Quality (STQ);
Speech samples and their use for QoS testing**

Reference
RTR/STQ-00215m

Keywords
QoS, quality, speech

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

IPRs essential or potentially essential to the present document 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.

Foreword

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

Modal verbs terminology

In the present document "**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

Conducting drive test in multi technology environment presents a challenge to all parties. And the complexity and variance of the different scenarios need to be broken down to handy instructions for those who actually configure and conduct the measurements, such as Network Operators, Service Providers, Equipment Vendors and Regulatory Authorities.

1 Scope

The present document introduces and explains the use and application of speech samples to determine the objective listening quality (LQO) in narrowband (NB), wideband (WB) and super-wideband (SWB) for different scenarios such as connections between fixed networks and mobile terminals.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

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 reference 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] Recommendation ITU-T P.48: "Specification for an intermediate reference system".

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

[i.3] Recommendation ITU-T P.830: "Subjective performance assessment of telephone-band and wideband digital codecs".

[i.4] Recommendation ITU-T P.862: "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".

[i.5] Recommendation ITU-T P.862.1: "Mapping function for transforming P.862 raw result scores to MOS-LQO".

[i.6] Recommendation ITU-T P.862.2: "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs".

[i.7] Recommendation ITU-T P.862.3: "Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2".

[i.8] Recommendation ITU-T P.863: "Perceptual objective listening quality assessment (POLQA)".

[i.9] Recommendation ITU-T P.863.1: "Application Guide for the Recommendation ITU-T P.863".

[i.10] Recommendation ITU-T G.711: "Pulse code modulation (PCM) of voice frequencies".

[i.11] Recommendation ITU-T G.191: "Software tools for speech and audio coding standardization".

[i.12] Recommendation ITU-T P.341: "Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals".

[i.13] Recommendation ITU-T P.56: "Objective measurement of active speech level".

[i.14] Recommendation ITU-T P.501: "Test signals for use in telephonometry".

3 Abbreviations

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

AMR	Adaptive Multi-Rate codec
AMR-WB	Adaptive Multi-Rate codec Wide Band
ASL	Active Speech Level
EFR	Enhance Full Rate codec
EVS	Enhanced Voice Services, speech codec
FIR	Finite Impulse Response filter
IRS	Intermediate Reference System
ISDN	Integrated Services Digital Network
LQO	Listening Quality Objective
MOS	Mean Opinion Score
MSIN	Mobile Station Input filter
NB	Narrow Band
NTP	Network Terminating Point
OVL	Overload point
PBX	Private Branch Exchange
PC	Personal Computer
PCM	Pulse Code Modulation
PSTN	Public Switch Telecommunication Network
SWB	Super Wide Band
VoLTE	Voice over LTE
WB	Wide Band

4 Devices and network access

4.1 Mobile devices

There are only a few devices and access interfaces that play a role in end-to-end mobile network testing. In end-to-end testing a test connection between two endpoints is established. This determines the access interfaces and devices.

The mobile device is not a pure access device to the mobile network. It contains complex components for speech processing and becomes therefore an important contributor to the overall quality measured in the test connection.

Mobile devices do not have a standardized audio interface, neither digital nor analogue. As common practice the headset connector of the mobile device is used as access interface for audio insertion and capturing. As a pre-condition for audio insertion and capturing, the measurement equipment has to match to the devices headset connector in impedance and level.

It has to be noted that in this setup the mobile devices are used in headset mode. Devices apply individual audio profiles, means individual settings in filtering, amplification and noise- and echo treatment for connected headphones or the use of the internal microphone. Often there is a third mode that applies when a handsfree loudspeaker set is connected. Since the audio processing in headphone mode is different from the use of internal microphone, such a test connection emulates a user with a headphone (personal handsfree kit) connected by wire to the headphone connector.

4.2 ISDN/PSTN

ISDN or (analogue) PSTN interfaces are not directly belonging to the mobile network but they are usually used as defined endpoint of the test connection. As access point to the ISDN or PSTN network a real consumer telephone device is not used but rather an ISDN or PSTN interface module as e.g. a PC card. It enables an electrical connection to the network for audio transmission and processes all the signalling information. The interface module or PC card is usually accessed with a digitalized speech signal in PCM format. The format is preferably 16 bit or 13 bit linear PCM sampled at 8 kHz or 16 kHz. Some interfaces expect 8 bit A-Law PCM that can be used in case of ISDN but is not recommended for PSTN, since it will cause an additional A-Law compression step in the test connection.

NOTE: The A-Law signal would be decompressed and fed as analogue signal in the local loop, where the regular A-Law compression will be at the digital NTP or the PBX.

Today, ISDN/PSTN channels are narrow-band only. Thus, a transmission to an ISDN/PSTN end-point is always restricted to narrowband despite that the airlink can use AMR-WB. The transition to narrowband is part of the gateway to the ISDN/PSTN.

4.3 Test scenarios

4.3.1 General aspects

The analogue circuits of almost all mobile devices are able to process wideband or fullband speech. Whether a call is transmitting narrowband or wideband or above speech depends on the wideband coding capability of the phone, the network and call setup. The subscriber cannot control whether the phone connects in narrowband, in wideband or in super wideband. The established channel determines the transmission bandwidth of the channel that can be narrowband, wideband, super-wideband or even fullband.

4.3.2 Narrowband telephony and narrowband test scenario

The conventional narrowband or normal-band telephony is traditionally using a pass-band from 300 Hz to 3 400 Hz. In digital transmission the technical limit is given by the Nyquist frequency due to sampling at 4 kHz upper audio transmission limit; there is no limit at the lower boundary. Today's narrowband speech codecs as EFR or AMR are also able to encode an audio band up to 4 kHz. Despite that fact, in practice a dedicated filtering is applied to the signal. Usually, there is a bandpass that is wider than the traditional pass-band but still limiting at the lower and upper range. The actual transmission characteristic is depending on the phone manufacturer and the setting of the phone. There are no binding limits or characteristics.

Testing narrowband is not tied to a narrowband channel. Narrowband testing means that the listening quality is estimated as listening through a conventional handset, the objective quality model filters the signal with such a band-pass and compares the speech signal to an ideal narrowband reference signal too. This restriction to a narrowband bandpass is applied despite the fact of the signal bandwidth passed through the channel.

For testing a narrowband scenario using a mobile access device there are two setups:

1) Insertion of a signal that exceeds the traditional narrowband bandwidth, e.g. 50 Hz to 3 800 Hz or even 50 Hz to 8 000 or 50 Hz to 14 000 Hz. In this case, the limitation of the signal is done by the device and the channel, while the device usually limits at most. At the receiving side, the recorded speech signal is compared to an ideal narrowband signal (at a bandwidth of 50 Hz to 3 800 Hz). In this test case the filter characteristic of the mobile device used has a significant influence on the estimated quality, since all restrictions to the reference bandwidth are considered as degradation. The predicted MOS describes the overall quality as it is perceived by the particular device and the channel; the score is device dependent.

2) Insertion of a signal that emulates a traditional sending path that is close to the defined passband of 300 Hz to 3 400 Hz. Therefore the test speech signal is filtered with a bandpass filter as e.g. IRSsend or MSIN. Usually, those filters are narrower than the phone's characteristic. The phone's band limitations will not affect significantly the speech signal anymore. By using such a setup, the filter characteristic of the particular phone becomes less influencing. The bandwidth of the signal at receiving side is than widely dominated by the applied pre-filtering and widely the same for all devices. The estimated score becomes less phone dependent.

The approach (1) is recommended for device testing. For field testing of mobile network quality the setup (2) is recommended. It focuses more on network quality than on device depending audio filtering.

Please note that the term narrowband test scenario does not depend on the actual transmission capability of the channel but rather on the quality reference that is just narrowband. Even a wideband channel can be tested in a narrowband setup, it can be compared to listening wideband with a traditional handset, the upper frequency ranges are just not perceptible by such a transducer.

Typical MOS scores in a narrowband scenario are:

- 4.5 for a complete transparent narrowband signal.
- 4.4 for an ISDN signal (coded with Recommendation ITU-T G.711 [i.10] A-Law).
- 4.2 to 4.3 for a perfectly processed signal with AMR at 12,2 kbit/s.
- 3.4 to 3.6 for a perfectly processed signal with AMR at 4,75 kbit/s.

Quality testing in a narrowband test scenario is used for a long time and most of published MOS scores relate to this scenario. The established Recommendation ITU-T P.862.1 [i.5] is an objective measure emulating a narrowband scenario. Also, the new Recommendation ITU-T P.863 [i.8] supports a dedicated narrowband test modus, where signal predictions are made according to a narrowband test setup.

4.3.3 Wideband telephony and super-wideband test scenario

For wideband telephony typically a transmission capability of 100 Hz to 7 000 Hz is defined. Similar to narrowband, the technical limits for a wideband transmission and channel are from often 50 Hz to 8 000 Hz due to the sampling frequency of 16 000 Hz.

NOTE: The AMR-WB speech codec limits at 6 400 Hz itself due to an internal sampling frequency of 12,8 kHz.

The step beyond wideband is called super-wideband and enables a transmission bandwidth from 50 Hz to 14 000 Hz. In practice, super-wideband can be seen as full-band for human speech, since there are no relevant signal parts above 14 000 Hz.

The recently standardized EVS speech codec supports all audio bandwidths from narrow-band, wideband, super-wideband and even to full-band. In comparison with AMR and AMR-WB, which can adapt bitrate but support only one fix audio bandwidths. The EVS speech codec can change both, audio bandwidth and bitrate and is able to choose to the best compromise between bitrate and audio bandwidth adaptively. For VoLTE the EVS codec will support super-wideband audio as default.

A wideband or super-wideband transmission requires a corresponding channel and two endpoint devices, those are able to process wideband or super-wideband speech. Today, wideband in the field can only be tested in mobile to mobile connections, since ISDN/PSTN are restricted to narrowband.

In a traditional wideband scenario, a wideband signal becomes compared to an ideal 100 Hz to 7 000 Hz or 50 Hz to 8 000 Hz signal. However, the there is a tendency to evaluate and score traditional wideband directly by comparing to super-wideband signal as an ideal reference. Along with the standardization of Recommendation ITU-T P.863 [i.8] there is the super-wideband mode recommended, where the recorded signal is compared with a 50 Hz to 14 000 Hz reference signal (Recommendation ITU-T P.862.2 [i.6] wideband supports a dedicated wideband modus, however this measure was not established in the field and superseded by Recommendation ITU-T P.863 [i.8] super-wideband mode).

The super-wideband scenario can be imagined as listening through a high quality headphone without perceptible restrictions in transmission. It is as a mono listening situation, where the same signal is perceived on both ears.

The actual limitation to 7 000 Hz or 8 000 Hz in a real wideband transmission as with the AMR-WB will lead to slight degradation compared to a reference of 50 Hz to 14 000 Hz. For testing a super-wideband channel, the wideband and super-wideband scenario is the best suited test scenario. In that scenario the signal can be evaluated completely up to its upper spectral range of 14 000 Hz. Super-wideband mode gives the possibility to relate each limitation to an ideal sample (quasi fullband reference) and can still be used in the future, when super-wideband codecs become deployed in mobile networks.

From a testing point of view, flat filtered super-wideband signal is inserted in the access interface. All limitations in bandwidth applied to the signal are taken into account. Typical MOS scores in a super-wideband scenario are:

- 4.75 for a full transparent signal from 50 Hz to 14 000 Hz or more.
- 4.2 to 4.5 for a full transparent wideband signal from 50 Hz to 7 000 or 8 000 Hz.
- 3.8 to 4.1 for a transparent processing with AMR-WB 12,65 and no further limitations in bandwidth.
- 3.2 to 3.5 for a transparent processing with AMR 12,2 in narrowband.

5 Speech samples

5.1 General aspects

Starting from the original speech sample recorded in the studio the sample need to be processed before they can be used in instrumental speech testing.

Speech samples for quality testing are usually composed by a subsequent series of sentences spoken by a human speaker. Traditionally, a sentence pair of two sentences is used in auditory tests following Recommendation ITU-T P.800 [i.2] and for instrumental testing as well.

Recommendations on recording and processing of speech samples for testing speech quality are given in Recommendations ITU-T P.800 [i.2] and P.830 [i.3]. Speech samples to be used for instrumental testing of speech quality have to fulfil additional technical requirements regarding temporal structure, noise floor and similar. Those recommendations are given in Recommendations ITU-T P.862.3 [i.7] and P.863.1 [i.9].

Typically, there is a systematic difference in scoring male or female voices, where male voices are scored by instrumental measures like P.862 [i.4] and P.863 [i.8]. For the purpose of automated testing as in drive test tools, speech samples combining sentences spoken by a male and a female talker is a preferable solution.

5.2 Pre-filtering of speech signals

5.2.1 Emulation of handsets

Depending on the application to be tested different filters need to be applied. In this context, filtering applies to an upfront filtering applied to the speech signal before it becomes inserted in the test device or the network interface respectively. This filter emulates the transmission characteristic of the microphone and its connection circuit, which is not present in an electrical insertion. After filtering, the signal becomes closer to the signal that would naturally be available at this point of insertion.

5.2.2 Filter for narrow-band test scenarios

5.2.2.1 IRS send Filter

The IRS filter (IRS stands for Intermediate Reference System) emulates a transmission characteristic of a traditional narrowband handset. There is an IRS send filter for the microphone and sending characteristic and an IRS receive filter for the characteristic of the receiving side including a (electro-dynamic) transducer.

The IRS send filter can be imagined as a bandfilter slightly wider than the normal passband but with a significant pre-emphasis towards 2 700 Hz. The classical IRS filters are defined in Recommendation ITU-T P.48 [i.1].

There is a revised characteristic (Modified IRS send) defined in Recommendation ITU-T P.830 [i.3] that has slightly weaker roll-off characteristics at the band limits. The difference at the upper boundary becomes much smaller, when a downsampling filter to 8 000 Hz is applied to the IRS filtered signal that is usual for input signals in a narrow-band channel (figure 1).

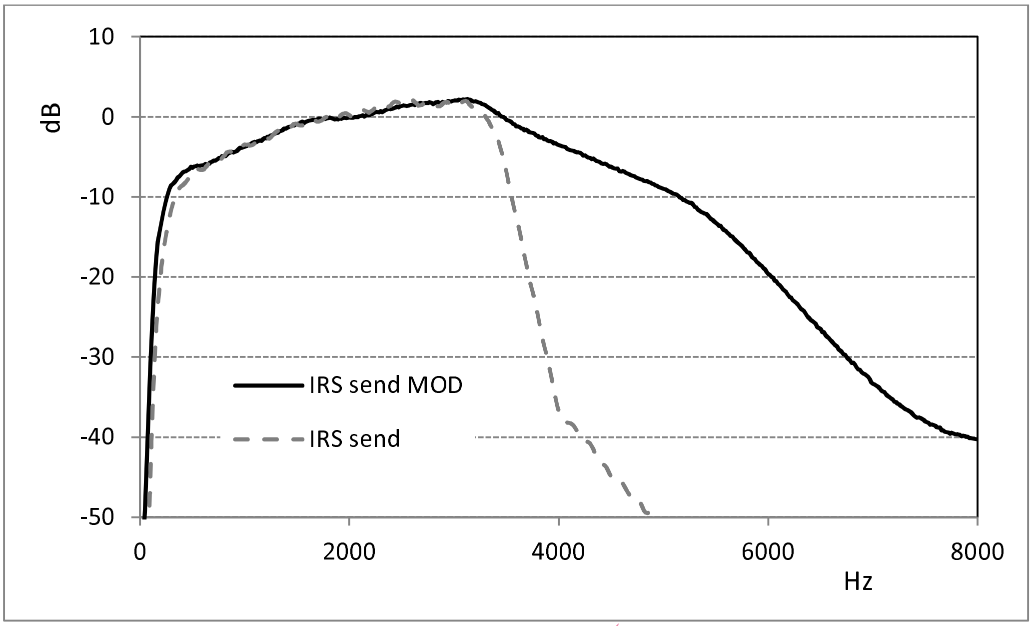

Figure 1: Frequency responses for IRS send and Modified IRS send filters

'Modified IRS send' is the pre-filter that is used by ITU-T for testing and evaluating narrowband speech codecs. IRS send and Modified IRS send filters are provided as examples in Recommendation ITU-T G.191 [i.11] that is a collection of processing algorithms of ITU-T.

5.2.2.2 MSIN Filter

The MSIN filter is also emulating a sending device but has no pre-emphasis (it is almost flat) and let pass lower frequencies compared to a Modified IRS send filter. MSIN is used in codec standardization too, but more related to cordless and mobile transmission components. The MSIN filter is also realized as an example implementation in Recommendation ITU-T G.191 [i.11].

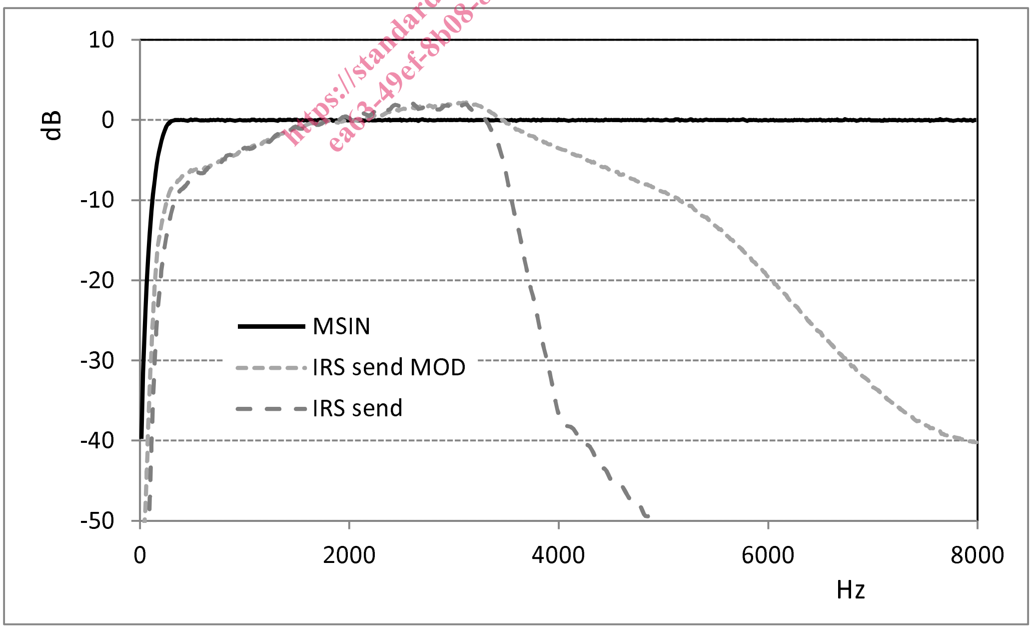

Figure 2: Frequency responses for IRS send and Modified IRS send filters compared to MSIN