ETSI TR 103 138 V1.6.1 (2023-03)

TECHNICAL REPORT

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

Reference

RTR/STQ-00228m

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

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

If you find a security vulnerability in the present document, please report it through our Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

BLUETOOTH® is a trademark registered and owned by Bluetooth SIG, Inc.

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), Super-Wideband (SWB) and Fullband (FB) for different scenarios such as connections between fixed networks and mobile terminals.

This revision of the present document reflects latest technologies and standards in voice transmission and evaluation.

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 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] 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] Void.

[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 prediction".

[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".

[i.15]	Recommendation ITU-T P.10/G100: "Vocabulary for performance, quality of service and quality of experience".
[i.16]	Recommendation ITU-T P.565.1: "Machine learning model for the assessment of transmission network impact on speech quality for mobile packet-switched voice services".
[i.17]	Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".
[i.18]	Recommendation ITU-T G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

3 Definition of terms, symbols and abbreviations

3.1 Terms

Void.

3.2 Symbols

Void.

3.3 Abbreviations

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

AMR	Adaptive Multi-Rate codec
AMR-WB	Adaptive Multi-Rate codec Wideband
ASL	Active Speech Level
EFR	Enhance Full Rate codec
EVS	Enhanced Voice Services, speech codec
FB	Fullband
FIR	Finite Impulse Response filter
IRAT call	Inter-Radio-Technology call
IR-HO	Inter-Radio-technology Handover
IRS	Intermediate Reference System
ISDN	Integrated Services Digital Network
LQO	Listening Quality Objective
MOS	Mean Opinion Score
MSIN	Mobile Station Input filter
NB	Narrowband
NTP	Network Terminating Point
OVL	Overload point
PBX	Private Branch Exchange
PC	Personal Computer
PCM	Pulse Code Modulation
PSTN	Public Switched Telephone Network
SRVCC	Single Radio Voice Call Continuity
SWB	Super-Wideband
VoLTE	Voice over LTE
WB	Wideband

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.

The user has to be aware that the test scenario is 'mobile device equipped with external headset' if the audio connector is being used for insertion and capturing speech signals. The applied speech processing and its effect on speech quality might be different from commonly used handset mode.

Smartphones as mobile devices partially offer internal interfaces for digital access to the voice signals. The main advantage is that the devices remains in handset mode and apply the related speech processing as in this model. Secondly, there is no additional A/D conversion needed and the risk of noise interference is minimized.

4.2 ISDN/PSTN interfaces and SIP landline clients

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 narrowband 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.

Landline devices can also be realized as SIP-based clients on a digital IP interface. SIP clients can operate using narrowband codecs as Recommendation ITU-T G.711 [i.10] or support wideband codecs as e.g. Recommendation ITU-T G.722 [i.17] or Recommendation ITU-T G.722.2 [i.18].

4.3 Test scenarios

4.3.1 General aspects

The term test scenario defines the context of the quality measurement in a subjective experiment or by an instrumental method. It is defined by the presented audio band width of the reference signal (see clause 5.5). This can be fullband (exceptionally: super-wideband), wideband or narrowband. The most common test scenario today is fullband, which means a signal after transmission and potentially limitation in audio bandwidth is rated against a fullband reference signal. The fullband test scenario gives the possibility to score limitations is bandwidth relatively to a perfect fullband and therefore allows the direct comparison of perceived degradations by bandwidth limitation.

NOTE 1: Because, those fullband test scenarios are also applied to wideband or narrow band terminals or transmission systems, they are also called 'mixed band scenario'.

The dedicated wideband test scenario, where the reference signal in the listening test or in an instrumental approach is a wideband signal (audio bandwidth limited at 8 kHz) was replaced by mixed or fullband scenarios in the last years.

The narrowband test scenario was widely used in the past before wideband transmission was deployed. Here the reference signal is a narrowband signal. A narrowband test scenario cannot be applied to wideband or full band signals and should be restricted to special cases like the comparison to historical results.

NOTE 2: There is no mean to compare or transform speech quality scores obtained in a narrowband scenario to quality scores obtained in a fullband scenario.

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 like 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.

The narrowband test scenario can be imagined as a listening situation, where a listener perceives the speech signal by a conventionally shaped handset. In this case, the transducer is limited to narrowband and also the listener does not expect a wider audio band than narrowband. A perfect, undistorted narrowband signal fulfils the expectation of a listener fully.

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 published MOS scores relate to this scenario. The Recommendation ITU-T P.863 [i.8] supports a dedicated narrowband test mode, where signal predictions are made according to a narrowband test setup. This test mode is mainly for dedicated use cases like e.g. the comparison to historical narrowband data and backward compatibility with Recommendation ITU-T P.862.1 [i.5], that is an objective measure emulating a narrowband scenario.

4.3.3 Fullband test scenario

In the past, there have been dedicated subjective experiments for narrowband and wideband, where only test conditions have been evaluated at the same audio bandwidth. Along with super-wideband and fullband transmission as well as real field conditions it also became also a focus to compare different transmission scenarios and evaluate the influence of audio bandwidth directly in one test. It means, that the bandwidth became a variable in the experiment. Those experimental setups are usually called 'mixed bandwidth' or simplified 'super-wideband' or 'fullband'. Main characteristic of these tests is the presence of a fullband (or super-wideband) signal in the experiment as best quality reference. Hence, all test conditions are rated in comparison to this reference and a limited bandwidth is usually rated as degradation. To minimize biases by over- or under representation of certain bandwidth, a minimum ratio of each bandwidth-class is recommended in the listening experiment. This test design with mixed bandwidths was used during the development and training of Recommendation ITU-T P.863 [i.8].

The fullband scenario can be imagined as listening through a high-quality headphone without perceptible restrictions in transmission. It is like a mono listening situation, where the same signal is perceived on both ears. The undistorted fullband speech signal determines the best quality received in such a test scenario and is scored highest and often > 4,7 MOS. Perceptible degradations in the spectrum caused e.g. by compression or noises but also due to band limitations as in wideband or narrowband channels will lead to decreased quality scores.

The actual limitation to 7 000 Hz or 8 000 Hz in a real wideband transmission as with the AMR-WB [i.18] will lead already to a degradation compared to a fullband or super-wideband reference. This degradation is caused by the coding artifacts of the AMR-WB codec as well due to the bandwidth limitation to ~7 000 Hz.

Note, for testing speech communication channels and networks - despite their technically transmitted audio bandwidth - the fullband test scenario is the best suited test scenario. In that scenario the signal can be evaluated completely up to its upper spectral range. Fullband mode gives the possibility to relate each limitation to an ideal sample (fullband reference). It also reflects today's listening equipment like headphones connected top mobile devices and the expectation of subscribers. A limitation in audio bandwidth is perceived as lowered quality and should be considered in the evaluation.

Typical MOS in a fullband scenario obtained with speech samples as in Recommendation ITU-T P.501 [i.14], Annex C are:

- 4,79 for the fullband reference.
- 4,77 for a full transparent signal from 50 Hz to 14 000 Hz or more.
- 4,4 to 4,8 for a transparent processing with EVS 24,4 (50 Hz to 16 000 Hz).
- 4,3 to 4,7 for a full transparent wideband signal from 50 Hz to 7 000 or 8 000 Hz.
- 3,9 to 4,2 for a transparent processing with AMR-WB 23,85 and no further limitations in bandwidth.
- 3,7 to 4,1 for a transparent processing with AMR-WB 12,65 and no further limitations in bandwidth.
- 3,0 to 3,5 for a transparent processing with AMR 12,2 in narrowband.

Note that the actually obtained MOS is depending on the used speech signal and its characteristics which may lead to more or less degrading artifacts due to filtering and compressing.

Even Recommendation ITU-T P.863 [i.8] supports a dedicated narrowband mode for historic compatibility, the recommended mode for all sorts of transmission scenarios is the fullband mode, where the recorded signal is compared with a fullband reference signal.

5 Speech samples and signal processing

5.1 Void

5.2 Void

5.3 Void

5.4 Creating speech samples for speech quality testing

As reference sample an unprocessed speech sample usually in fullband is considered. This reference sample is used in the subjective experiment to mark the highest quality in the experiment. This reference signal also defines the undistorted reference signal required by so-called full-reference objective speech quality measures as Recommendation ITU-T P.863 [i.8] or as input signal for Recommendation ITU-T P.565.1 [i.16].

Starting from the original speech sample recorded in the studio, the sample need to be processed before they can be used for listening tests or in instrumental speech quality testing.

Speech samples for quality testing are usually composed of 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 Recommendation ITU-T P.800 [i.2] and Recommendation ITU-T 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 for speech samples 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 degraded male or female voices. This difference is observable in subjective listening tests and depends on coding schemes or other processing like audio band limitations. A systematic difference is also visible by instrumental measures like Recommendation ITU-T P.863 [i.8] and the previous Recommendation ITU-T P.862 [i.4]. For the purpose of automated testing as in drive test tools, speech samples combining sentences spoken by a male and a female talker is the preferable solution e.g. as given in Recommendation ITU-T P.501 [i.14], Annex D.

5.5 Reference signals for speech quality measurements

5.5.1 Reference signals for testing in fullband test scenarios

The term reference signal is used in two contexts. In a subjective listening test, the reference signal usually defines the highest achievable quality with an undistorted signal that has no audio band limitation. Furthermore, the term reference signal is also used for instrumental, objective full-reference measures, which require a reference signal for comparison with the degraded signal. Example is the Recommendation ITU-T P.863 [i.8].

While the input signal into the terminal or transmission system might be pre-filtered and adapted to the access device, the reference signal remains largely unprocessed and has to preserve the minimum spectral range for the test scenario.

The fullband test scenario is the recommended and most common scenario today.

NOTE: The term 'Test scenario' defines the highest audio bandwidth where signals are rated to. It is not identical with the actual transmission bandwidth of the terminal or system (see clause 4.3.3).

Therefore, fullband reference signals are also used for testing wideband and narrowband transmission systems. A fullband signal is recommended as reference signal in an fullband or a mixed subjective test setup as well as a reference signal to Recommendation ITU-T P.863 [i.8] in the default fullband mode. Any limitation to this bandwidth is considered a degradation in a fullband test scenario.

For tests with Recommendation ITU-T P.863 [i.8] in Fullband mode (FB), the reference signal has to be a flat signal that is unfiltered and exceptionally may have a high-pass at 20 Hz or 50 Hz to remove very low-frequency noises. There should be no further low-pass <20 kHz be applied to the signal. In fullband mode testing, the reference signal is typically the same as the input signals described in clause 5.6.2.1.

5.5.2 Reference signals for testing in narrowband scenario

Reference signals applicable for narrowband scenarios are limited at 4 kHz audio bandwidth. To receive such narrowband reference signals, the original signals have to be reduced in sampling rate and bandwidth. With the reconstruction filter as realized in Recommendation ITU-T G.191 [i.11] which applies a cut-off at 0,9 of the Nyquist frequency, the required bandwidth of 3 800 Hz with a sampling frequency of 8 kHz cannot be guaranteed. The lowpass filter given in Annex A overcomes that shortcoming.

Recommendation ITU-T P.863 [i.8] supports a dedicated narrowband mode for historic compatibility with e.g. Recommendation ITU-T P.862 [i.4]. For tests in the NB mode a reference signal is preferred with 8 kHz but also 48 kHz sampling frequency is allowed. In either case a minimum bandwidth of up to 3 800 Hz in required. In practice, reference files for Recommendation ITU-T P.863 [i.8] NB measurements are sampled with 8 kHz which also provides a backwards compatibility to Recommendation ITU-T P.862.1 [i.5], where a reference signal sampled with 8 kHz is recommended too.

5.6 Pre-processing input speech signals for transmission

5.6.1 General aspects

Terminals and transmission systems have dedicated requirements on inserting speech signals caused by the interfaces. Those requirements usually are:

- Sampling frequency (in case of digital insertion).
- Minimum audio bandwidth.
- Potential pre-filtering of signals.
- Speech signal level.

In most of the test cases, flat filtered input speech signals up to a minimum audio bandwidth are recommended. However, in special cases a terminal emulation by pre-filtering is appreciated.

5.6.2 Pre-processing for super-wideband and fullband transmission

5.6.2.1 Speech signals for super-wideband and fullband telephony

Speech signals for super-wideband and fullband terminals and transmission systems should cover an audio bandwidth of 20 kHz or above. The corresponding sampling frequency is 48 kHz. If required by the terminal or transmission system interface, a sampling frequency of 44,1 kHz can be applied too. Fullband input signals are also used as input into wideband and narrowband transmission systems.

5.6.2.2 Filter for fullband speech signals

As a filter for the fullband speech, a bandpass from 20 Hz to 20 000 Hz with a flat bandpass can be applied to an unfiltered reference signal exceeding this bandwidth. A reference is given in Recommendation ITU-T P.10/G.100 [i.15].

5.6.2.3 Application of 14 kHz bandpass

As a filter for the so-called super-wideband speech, often a bandpass from 50 Hz to 14 000 Hz with a flat bandpass was applied. A reference implementation is a filter described as '14KBP' in Recommendation ITU-T G.191 [i.11], that is the audio processing tool collection of ITU-T. The band limitation at 14 kHz was recommended for previous versions of Recommendation ITU-T P.863 [i.8] limited to super-wideband. Since edition 3 of the Recommendation ITU-T P.863 [i.8] was approved in 2018 the use of fullband audio as reference speech signals is recommended and the fullband analysis mode supersedes the previous super-wideband analysis mode completely. Note that the commonly used EVS 24,4 codec supports and transmits an audio spectrum up to 16 000 Hz and the use of a 14 kHz pre-filtered signal will not cover its full audio spectrum.

5.6.3 Pre-processing for wideband and narrowband transmission

5.6.3.1 Input speech signals for wideband and narrowband telephony

Speech signals for wideband or narrowband transmission need to cover at least the entire audio spectrum of the transmission system that is 4 kHz audio bandwidth for narrowband or 8 kHz for wideband telephony. However, in practice, usually fullband signals are used as input signals into wideband and narrowband systems, a limitation of bandwidth is applied by the terminal or the transmission system itself.

5.6.3.2 Filter for re-sampling wideband and narrowband signals

There can be interfaces to the terminal or the transmission system requiring a lower sampling frequency like e.g. 16 kHz or 8 kHz. In those cases, the input speech signal is re-sampled by applying a corresponding lowpass filter as in Annex A. In practice, the upper frequency is limited at 3 900 Hz for 8 kHz sampling frequency or 6 800 Hz.

NOTE: The high quality down sampling routine in Recommendation ITU-T G.191 [i.11] applies a reconstruction lowpass at 0,9 of the Nyquist frequency, therefore the signal sampled at 8 kHz will have a cut-off frequency of 3 600 Hz that might be too low for some test cases. See Annex A for an example of an improved filter.

5.6.4 Pre-filtering of input speech signals to emulate handsets

5.6.4.1 Overview

Depending on the application to be tested different filters can be applied. Those filters emulate typical terminal behaviour in sending direction as defined in Recommendation ITU-T P.48 [i.1] or Recommendation ITU-T P.830 [i.3]. In case, the speech signal is inserted into the device behind its internal filter or in a simulated context, these filters emulate the terminal send characteristic.

In this context, filtering applies to an upfront filtering applied to the speech signal before it is 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 or digital insertion. After filtering, the signal becomes closer to the signal that would naturally be available at this point of insertion in case the terminal applies filters and/or a pre-emphasis.

NOTE: The emulation of a traditional handset device characteristic - especially in sending direction - was commonly applied in the past for narrowband tests, where a electrical or digital input into the transmission system was used, like e.g. a PSTN or an ISDN card.

Smartphone devices have a flat input characteristics despite the supported bandwidth even for narrowband transmission. Therefore, the dedicated emulation of a traditional handset is not recommended for tests in fullband scenarios. In narrowband test scenarios and especially for historic comparisons, a handset emulation can be applied to the input signal to a smartphone.

In general, handset emulation is not recommended anymore in speech quality field testing.

5.6.4.2 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 band filter 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 down sampling filter to 8 000 Hz is applied to the IRS filtered signal which is common for input signals in a narrowband channel (see figure 1).

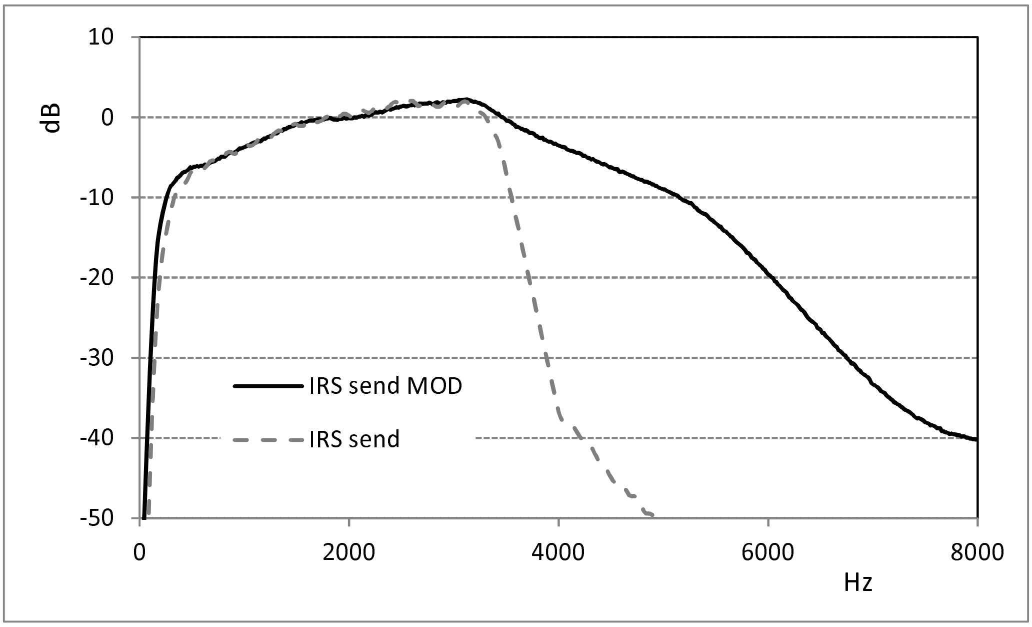

Figure 1: Frequency responses for IRS send and Modified IRS send filters

'Modified IRS send' is the pre-filter which 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] which is a collection of processing algorithms of ITU-T.

5.6.4.3 MSIN Filter

The MSIN filter is also emulating a sending device but has no pre-emphasis (it is almost flat) and allows for lower frequencies to pass 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].

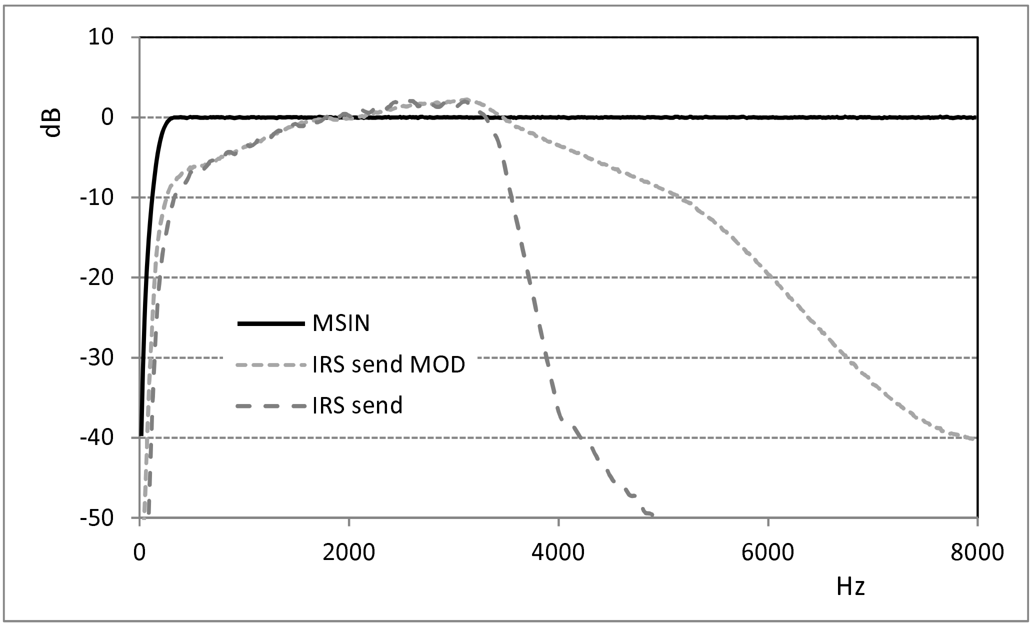

Figure 2: Frequency responses for IRS send and Modified IRS send filters compared to MSIN

5.6.4.4 Wideband filter according to Recommendation ITU-T P.341

Recommendation ITU-T P.341 [i.12] describes test setups for telephone devices and also a tolerance scheme for a common wideband filter. It is in principle a flat filter that filters wideband signals from 50 Hz to 7 000 Hz. The tolerance scheme in Recommendation ITU-T P.341 [i.12] allows a wide range of individual realizations of such a filter. In Recommendation ITU-T G.191 [i.11] a reference implementation is given and it is recommended to use this, in case a Recommendation ITU-T P.341 [i.12] filter needs to be applied.

Recommendation ITU-T P.341 [i.12] filters for wideband are - if at all - today used in offline processing and codec standardization and less in real field testing. The use of a Recommendation ITU-T P.341 [i.12] type filter upfront to the insertion of the speech signal is not recommended in case of insertion into a real device's headphone connector, but rather a super-wideband signal should be inserted in a wideband test case. A band- or low-pass will be applied by the device itself.

5.6.4.5 Compensation filters for analogue inputs

There can be interfaces to a device, where the signal is processed by a proprietary filter of the device. Those filters can be implemented e.g. to correct non-flat frequency responses of the microphone or the analogue circuits. In this case, an inserted signal at this interface will result in an unintended (e.g. non-flat) signal at the internal reference point of the device (e.g. encoder input), a compensation filter can be applied to compensate the proprietary filter of the device's input interface.

NOTE: Smartphone devices available at the time of publication of the present document do not apply those proprietary filters anymore, also analogue input interfaces transmit with flat frequency response. Exception can be soft high-pass filters to reduce low-frequency environmental noises.

5.6.5 Audio level

5.6.5.1 Nominal level

The signals to be inserted have to be scaled to a defined audio level for compatibility and to match the working range of the codec. In principle, a level that is too low will lead to a low S/N ratio with a noise floor of the analogue circuits and the quantization noise, a level that is too high may lead to amplitude clipping.

There is a nominal level at digital lines that is -26 dB OVL. In principle this level corresponds to -20 dBm at a 600 Ohms as for narrowband four-wire analogue interface. All speech codecs in telephony are optimized and tested for speech signals at this nominal level.

5.6.5.2 Level adjustment with Recommendation ITU-T P.56

Speech is a temporally fluctuating signal with pauses. Recommendations about the temporal structure related to active speech and pauses are given in Recommendation ITU-T P.862.3 [i.7] and Recommendation ITU-T P.863.1 [i.9].

The term Active Speech Level (ASL) refers to the rms level of the active speech parts only. ITU-T defines an algorithm for ASL measurements in the Recommendation ITU-T P.56 [i.13] 'Speech Voltmeter'. A speech sample at nominal level is normalized to -26 dB OVL according to Recommendation ITU-T P.56 [i.13].

5.6.5.3 Input level at test devices

A speech signal at nominal level of -26 dB OVL can directly be used on an ISDN/PSTN interface. It is the direct linear transition to the channel, where -26 dB OVL applies as nominal level.

The correct adjustment of the audio level at the microphone input of the mobile device is more critical. The microphone path usually applies a strong gain for low level microphone signals. Therefore, the speech signal has to be attenuated accordingly. The inserted speech signal has to be attenuated to a value that is transformed to -26 dB OVL at the input of the codec in the mobile device.

5.7 Devices and networks

5.7.1 Overview

The analogue circuits of almost all mobile devices are able to process fullband speech. However, the actual bandwidth transported to a far-end side in telephony depends on the audio 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 audio bandwidth of the channel that can be narrowband, wideband, super-wideband or even fullband. The audio bandwidth and the coding scheme can even change within an ongoing call. There can be cases where different limitations in audio bandwidths are observable in one speech sample.

A wideband, super-wideband or fullband transmission needs a corresponding channel and two endpoint devices, that are able to process wideband, super-wideband or fullband speech. At the time of publication of the present document, wideband or super-wideband transmission in the field can only be tested in mobile to mobile connections, since ISDN/PSTN are restricted to narrowband.

5.7.2 Insertion of speech signals into the device

5.7.2.1 Insertion at a digital interface

The prepared input speech signals have to be inserted into the device. Landline interfaces like PSTN- or ISDN-cards usually provide a digital interface, where the signal can be inserted with a sampling frequency of 8 kHz or 16 kHz at nominal speech level. SIP landlines are usually operating at 16 kHz sampling frequency in case of digital insertion.

Insertion into a mobile phone can be done either directly on the phone at a digital audio interface or by connecting an external play-out device at the analogue headset connector.

The insertion at a digital interface allows the use of the mobile device in handset mode, the device applies the same audio processing as for using of the internal microphone. The used digital interface determines the sampling frequency to be used which typically is 48 kHz, but there can be individual interfaces to dedicated services limited to narrowband or wideband transmission requiring a sampling frequency of 16 kHz. Typically, a digital interface can be directly supplied by signals at nominal level (see clause 5.6.5.1).

5.7.2.2 Precaution for insertion at an analogue interface

The insertion of speech signals into the external microphone input at the headset connector requires an external play-out device preferably operating at 48 kHz sampling frequency without any band limitation within the limits of fullband audio (20 Hz to 20 000 Hz).

NOTE: In case the external microphone input is connected to the headset connector, the mobile devices are switching in headset mode. Devices may usually apply different audio profiles, means individual settings in filtering, amplification and noise- and echo treatment for connected headphones compared to handset mode, where the internal microphone and loudspeaker are used (see clause 4.1).

The microphone input is designed for very low input levels as provided by external microphones. The actual analogue levels that need to be provided are a few millivolts only. The levelling has to be adjusted to achieve nominal level at the encoder input of the device or its internal digital interface. These very low input levels makes this type of insertion prone to interferences to the cable used and/or non-optimal grounding.

This level adjustment is not trivial if there is no digital interface is available on the phone itself. Therefore, digital insertion is to prefer to analogue insertion into the headset connector.

Some mobile devices enable other audio interfaces like Bluetooth® or USB. It has to be considered that Bluetooth will apply an additional coding. In both cases, the device applies audio processing as implemented for external Bluetooth or USB headsets. Hence, inserting at those interfaces, means a use case where the signal represents a connected Bluetooth device or an USB headset. Potential effects and degradations are considered in the quality evaluation.

5.7.3 Narrowband devices and networks

Narrowband telephony has been the standard for decades and applies also today if setting up channels to land-line devices, in interoperator calls or as fallback in case a certain device is not supported in wideband or above by a network.

Classical narrowband with a sharp limitation to 300 Hz to 3 400 Hz passband is not used anymore today. Transmission in ISDN or PSTN networks support an audio bandwidth up to almost 4 000 Hz. There is a limitation in case of digital transport and its sampling frequency of 8 kHz in narrowband. In digital narrowband networks the Recommendation ITU-T G.711 [i.10] codec is used, either in A-law or on µ-Law mode.

Typically, the AMR codec is used in mobile connections. This codec can apply different bitrates from 4,75 kbit/s up to 12,2 kbit/s, where the 12,2 kbit/s mode is identical to the former GSM-EFR codec. The AMR codec does not use the entire audio bandwidth, it ends at around 3 600 Hz due to internal processing schemes.

NOTE: Typically, in an established mobile narrowband channel both devices use AMR. However, there are cases, where a mobile device uses AMR-WB or even EVS in VoLTE and the network limits the actual audio bandwidth due to transcoding, e.g. by Recommendation ITU-T G.711 [i.10] for network internal transport or as fallback to minimum capabilities in case of device issues.

As input signal into a narrowband channel or system, the audio bandwidth needs to cover at least 4 000 Hz. In practice, a fullband signal is used in case of an analogue input or a sufficient sampling frequency for digital insertion. In case the digital interface does support only 16 kHz or 8 kHz (e.g. ISDN cards), the signal can be downsampled under the constraint of an upper bandwidth boundary close to 4 000 Hz. An example for appropriate sampling lowpass filters is given in Annex A.

5.7.4 Wideband devices and networks

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 channel are from often 50 Hz to 8 000 Hz due to the sampling frequency of 16 000 Hz. Landlines connected to SIP clients supports usually wideband coding with a sampling frequency of 16 kHz.

NOTE 1: The AMR-WB speech codec limits itself at 6 400 Hz due to an internal sampling frequency of 12,8 kHz.

NOTE 2: The EVS speech codec in wideband mode as well as many proprietary codecs encode an audio bandwidth up to 8 000 Hz.

NOTE 3: The EVS AMRWB-IO decoder reproduces a speech signal up to 8 000 Hz if decoding a standard encoded AMR-WB bitstream. The upper part of the signal that exceeds the AMR-WB capabilities is extrapolated artificially by the EVS AMRWB-IO decoder.

As input signal into a wideband channel or system, the audio bandwidth needs to cover at least 8 000 Hz. In practice, a fullband signal is used in case of an analogue input and a sufficient sampling frequency for digital insertion.

In case the digital interface does support only 16 kHz, the signal can be downsamples under the constraint of an upper bandwidth boundary close to 8 000 Hz. An example for appropriate sampling lowpass filters is given in Annex A.

5.7.5 Super-wideband devices and networks

The next step beyond wideband is called super-wideband and enables a transmission bandwidth from 50 Hz to 14 000 Hz or 50 Hz to 16 000 Hz. In practice, super-wideband can be seen as equivalent to fullband for human speech, since there are no relevant signal parts in speech above 14 000 Hz.

The AMR and AMR-WB codecs can adapt the bitrate but support only one fix audio bandwidth. In comparison, the EVS speech codec supports all audio bandwidths from narrowband, wideband, super-wideband and even to full-band. It 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.

NOTE: The EVS codec in SWB mode supports an audio bandwidth up to 16 kHz at the typical 24,4 kbit/s bitrate. The EVS in SWB mode at 13,2 kbit/s, as used in some regions, supports an audio bandwidth up to 14 kHz. Both bandwidth are named 'super-wideband'.

From a testing point of view, a fullband signal or exceptionally a flat filtered super-wideband is inserted in the access interface. All limitations in bandwidth applied to the signal are considered.

5.7.6 IRAT and SRVCC scenarios

Device and network capabilities determine whether a channel is narrowband, wideband or super-wideband. Usually, a channel is established as a narrowband, wideband or super-wideband from end to end.

However, voice calls can be setup in between devices being registered in different technologies applying different coding schemes (IRAT calls). In those scenarios, a transcoding of the speech signals in the network is necessary to match the supported codecs at either side. Because, transcoding is often linked to changes in bandwidth (e.g. EVS super-wideband in VoLTE to AMR-WB in wideband or a wideband or super-wideband mobile channel to a narrowband landline). In those cases, in one part of the channel a different coding scheme and a different audio bandwidth can be supported. Hence, the lowest bandwidth will be perceptually dominating.

Technology changes are not static, rather can happen dynamically at any time. In case of changing network conditions a handover to another technology can even happen during an ongoing voice call (IR-HO, SRVCC). In those cases, the change of the coding scheme happens without releasing the call. As a consequence there can be calls where the individual voice samples are processed by different codecs and audio bandwidths. Those changes can even happen during active speech and lead to a change of codec and potentially bandwidth within a single speech sample.

5.8 Capturing transmitted speech signals

5.8.1 Capturing speech signals from device

Automated speech quality prediction are usually analyses the transmitted and captured and potentially degraded audio signals, according to measures Recommendation ITU-T P.863 [i.8] or Recommendation ITU-T P.862.1 [i.5]. The captured audio signals can also be used for subjective listening tests. Capturing speech signals will not apply to quality evaluation approaches based on other information than audio as e.g. Recommendation ITU-T P.565.1 [i.16].

Capturing from landline interfaces like PSTN- or ISDN- cards is usually based on direct digital access and the audio signal is provided as decoded signal with 16 kHz or 8 kHz sampling frequency. Those signals are preferably re-sampled to 48 kHz before used as input signals to Recommendation ITU-T P.863 [i.8] in fullband mode.

Capturing on mobile phones is done either directly on the phone at a digital audio interface or by connecting a recording device at the analogue headset connector.

The recording at a digital interface allows capturing in handset mode, the device applies the same audio processing as for using the internal loudspeaker. If recording the decoded speech signal at a digital interface, the sampling frequency is usually 48 kHz, but there can be individual interfaces to dedicated services limited to narrowband or wideband transmission offering 16 kHz as sampling frequency. In this case, a separate re-sampling to 48 kHz is recommended before the speech signal is used as input signals to Recommendation ITU-T P.863 [i.8] in fullband mode.

The recording of decoded speech signals at the headset connector requires an external recording device preferably operating at 48 kHz sampling frequency, at least 16 bit quantization and without any band limitation within the limits of fullband audio (20 Hz to 20 000 Hz).

NOTE: In case of connecting an external recording device to the headset connector, the mobile devices are used in headset mode. Devices may apply a different audio profiles, means individual settings in filtering, amplification and noise- and echo treatment for connected headphones compared to handset mode, where internal microphone and loudspeaker are used (see clause 4.1).

Some mobile devices enable other audio interfaces like Bluetooth or USB. It has to be considered that Bluetooth will apply an additional coding. In both cases, the device applies audio processing as implemented for external Bluetooth and USB headsets. Hence, inserting at those interfaces means testing a use case where the signal represents a connected Bluetooth device or a USB headset. Potential effects and degradations are considered in the quality evaluation.

5.8.2 Compensation filters

In the past, narrowband test scenarios signals captured digitally at ISDN- or PSTN interfaces have regularly been post-filtered by an IRS Receive filter emulating the play-out frequency response of a traditional handset device. This filter is not applied anymore today.

The audio output from mobile devices is usually flat filtered today and there are no considerable limitations in bandwidth. However, in rare cases there can be non-flat filter characteristics applied by the phone to compensate for the internal loudspeaker frequency response. There is the option to equalize those frequency responses by a corresponding post-filter step to achieve a flat frequency response in the signal before speech quality evaluation. However, those compensation filters should be used with care. They should not be used in case the frequency response of the device does not affect the speech quality prediction negatively.

5.8.3 Requirements for automated quality prediction of speech signals

Speech-signal based objective models for automated speech quality prediction can be separated in non-reference and full-reference approaches. While the full-reference approach compares the degraded signal to an undistorted reference signal, the non-reference models predict the quality solely on analysis of the degraded signal.

Independent from the model approach, the captured speech signals should fulfil some requirements to be analysed by the quality prediction models. Typically, the models require a given digital format, like a 16 bit linear PCM signal at a certain sampling frequency, for example Recommendation ITU-T P.863 [i.8] requires the signal in mono wave format or raw PCM and preferably samples at 48 kHz sampling frequency. For using the narrowband mode of Recommendation ITU-T P.863 [i.8], the signals can also be sampled at 8 kHz.

To avoid unnecessary re-sampling, the capturing interface should already provide signals in the required format.

Especially, as analogue capturing interfaces are uncalibrated, the signal level depends on volume settings and connector circuits. Usually, automated speech quality measurements in field-tests are based on speech signals on nominal level (see clause 5.6.5.1). A level correction to -26 dB OVL (speech level according to Recommendation ITU-T P.56 [i.13]) is recommended.

NOTE 1: Recommendation ITU-T P.863 [i.8] in fullband mode considers a non-optimal level as quality degradation, an arbitrary level differing from nominal level may lead to unwanted effects on speech quality scores.

Speech quality measurements also depend on the amount of noise in the speech signal. In case of a speech signal having a considerable noise floor, the length of the periods in the signal with noise but without speech have an influence on speech quality. The ratio of active speech to noisy pauses should be constant in case the quality values have to be compared afterwards. Best practice is a leading and trailing pause in between of 0,5 s to 1 s duration each before and after speech activity.

NOTE 2: In case of using of Recommendation ITU-T P.863 [i.8] a leading and trailing pause of 0,5 s to 1 s is recommended, independent of a potential noise floor. Likewise, a pause in between the typical sentence pairs of about 0,5 s is needed. The algorithms requires periods without speech activity to adjust the internal voice activity detector. Furthermore, the captured signals for evaluation should have a length of 6 s to 12 s in length an should contain at least 3,2 s of active speech (see Recommendation ITU-T P.863.1 [i.9]).

It is recommended to insert and to capture the signals in fullband audio. For a potential evaluation in narrowband mode the signals can be re-sampled to 8 kHz in a post-processing step by applying a high quality down-sampling routine. See Annex A for an example of an appropriate reconstruction lowpass filter.

Speech signals processed according to the recommendations above can also used in subjective listening tests to assess speech quality.

6 Example measurement setups

6.1 Void

6.2 Void

6.3 Void

6.4 Fullband mobile to mobile measurement

The mobile to mobile measurement setup in fullband mode is the most common and recommended default test scenario today. It can be applied to all technologies and transmission audio bandwidths. The reference speech sample has to fulfil the requirements for fullband signals.

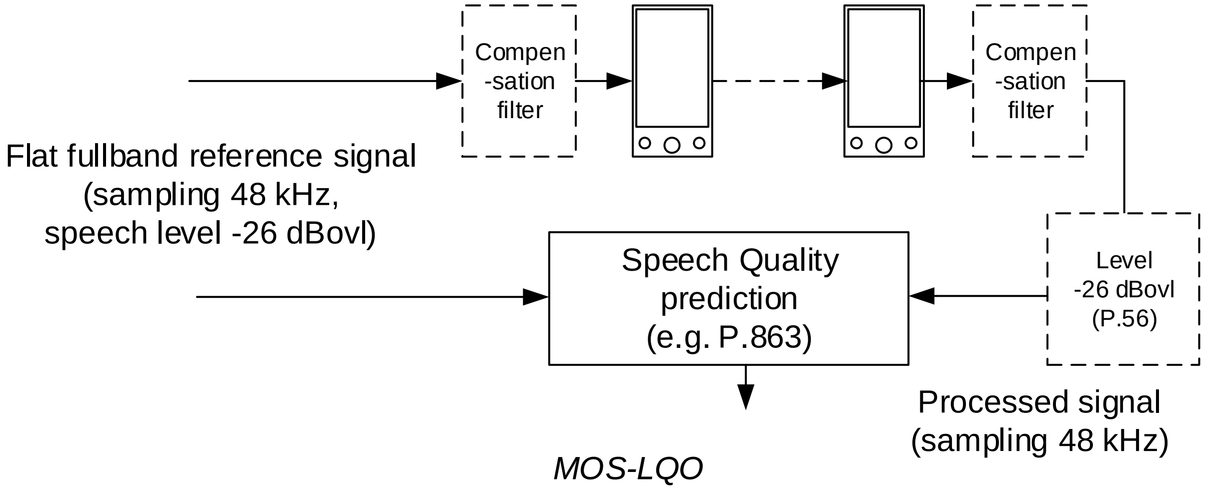
Compen-sation filter
Compen-sation filter
Flat fullband reference signal (sampling 48 kHz, speech level -26 dBovl)
Speech Quality prediction (e.g. P.863)
Level -26 dBovl (P.56)
Processed signal (sampling 48 kHz)
MOS-LQO

Figure 3: Test setup mobile to mobile

6.5 Fullband mobile to landline measurement

Although landline interfaces usually support only narrowband, a fullband test setup is recommended. Also in this case, the reference speech sample has to fulfil the requirements for fullband signals.

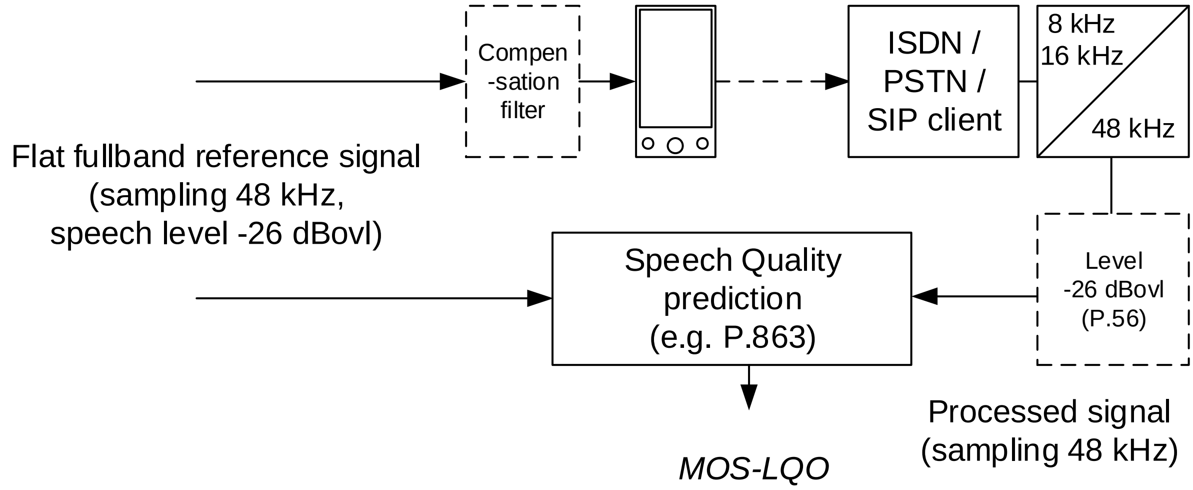

Figure 4: Test setup mobile to landline

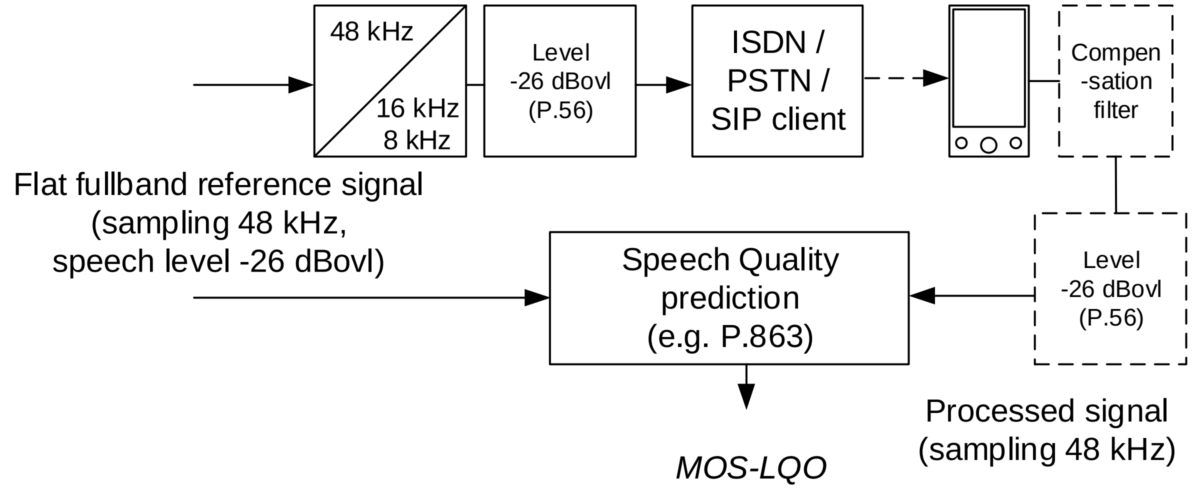

Figure 5: Test setup landline to mobile

7 Void

Annex A: Coefficients for the reconstruction lowpass filter

The coefficients for reconstruction lowpass filters ("ResampCoeff.h" in C/C++ notation) are contained in archive tr_103138v010601p0.zip which accompanies the present document.

The filters are designed for a cut-off frequency close to 0,95 of the Nyquist frequency and therefore allow a flat response up to 3 800 Hz when sampled at 8 kHz. There are coefficients for an up- and downsampling by factor two, three and four. Please consider the right length of the FIR-filter for up- and down sampling to have the same steepness of the filter in all cases. The filters are designed as linear-phase FIR filters with a group delay of half the filter length. A constant set of filters and these coefficients may be used for all kind of up-and downsampling.

Annex B: Bibliography

Void.

Annex C:
Speech Samples

C.1 Introduction

In the following a set of speech samples in different languages are presented those meet the requirements as described in the present document. The samples are based on speech material published in Recommendation ITU-T P.501 [i.14].

NOTE: The provided sample in British English is not based on Recommendation ITU-T P.501 [i.14] speech material rather composed of speech material used in the evaluation of Recommendation ITU-T P.863 [i.8].

C.2 Design

Recommendation ITU-T P.501 [i.14] provides 32 sentences spoken in eight languages by two male and two female talkers. These 32 sentences follow the technical specification as given in Recommendation ITU-T P.863.1 [i.9] and Recommendation ITU-T P.862.3 [i.7]. Subjective and objective scores obtained for a given scenario depend also on the speech sample, and more on the talker and gender. This leads to systematic differences in quality scoring depending on the used speech sample.

To minimize the gender dependency, speech samples can be composed of male and female talk spurts. Especially for mobile field testing sentence pairs consisting of one male and one female sentence are commonly used in practice. This Annex provides a set of composed sentence pairs in different languages.

For the following languages a composed speech sample is provided:

• American English	P501_D_AM_fm
• Chinese (Mandarin)	P501_D_CN_fm
• Dutch	P501_D_DU_fm
• British English	P501_D_EN_fm
• German	P501_D_GE_fm
• Finnish	P501_D_FI_fm
• French	P501_D_FR_fm
• Italian	P501_D_IT_fm

Each of these male/female composed samples balances the systematic bias between male and female voices as known for Recommendation ITU-T P.862 [i.4] and Recommendation ITU-T P.863 [i.8]. Additionally, the sentences and talkers have been selected to match MOS predictions for typical codec conditions that can be observed as averages over larger sets of speech samples. The procedure of processing and scheme of presentation follows exactly the way of presentation in Recommendation ITU-T P.863.1 [i.9].

Each sample is 6 s in length, it has a leading and a trailing pause as well as a pause in between the two sentences that meets the requirements in Recommendation ITU-T P.863.1 [i.9] and Recommendation ITU-T P.862.3 [i.7]. The noise floor in the speech pauses is < 85 dB (A) rmse but not digital silence.

C.3 Example results

As the result of a careful selection, all these samples show a much smaller language specific bias than an arbitrary chosen sample from Recommendation ITU-T P.501 [i.14]. Each individual sample matches well the average of many languages as represented by the average of all 32 speech samples in Recommendation ITU-T P.501 [i.14]. The deviation of the provided composed speech samples is much lesser than the original speech samples in Recommendation ITU-T P.501 [i.14].

Finally, all introduced samples in this annex can be considered as fully transparent as recommended for Recommendation ITU-T P.863 [i.8].

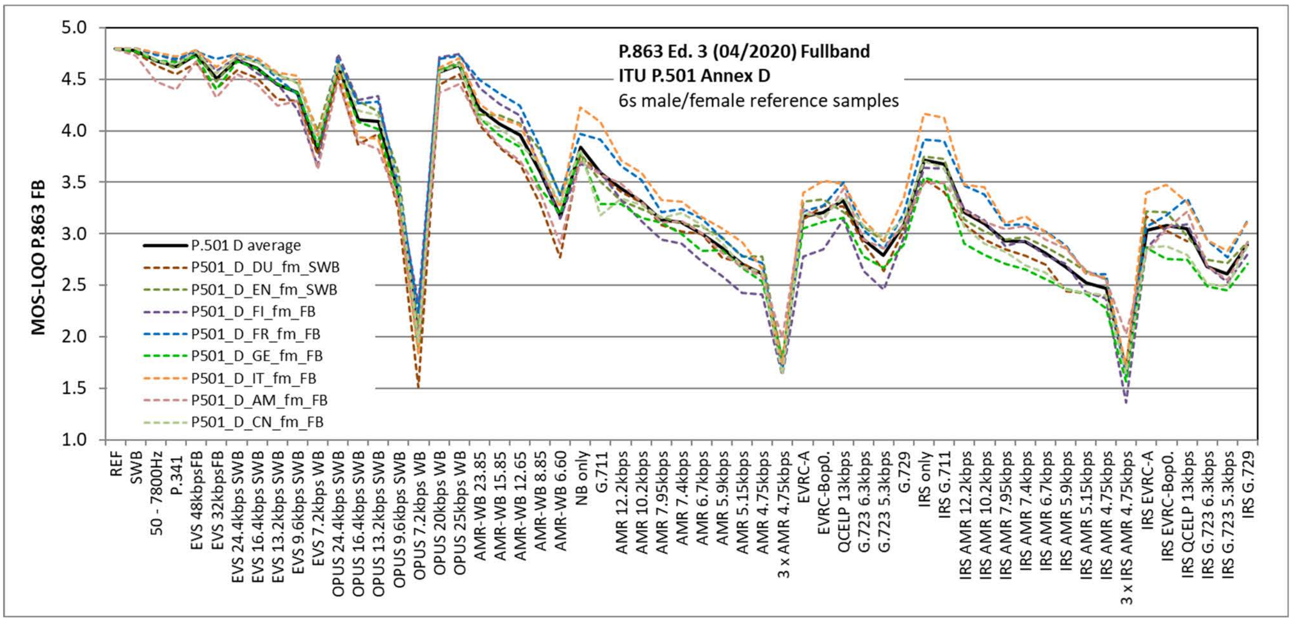

Figure C.1: Recommendation ITU-T P.863 [i.8] SWB example results of the composed speech samples in comparison to the average of all 32 samples from Recommendation ITU-T P.501 [i.14] for specific codec processing conditions

A corresponding analysis for Recommendation ITU-T P.863 [i.8] in narrowband mode shows a very good correspondence too.

C.4 Technical specification

To meet the recommendations as given in clauses 5 and 6 of the present document, all samples are provided in different sampling rates and pre-filters as given in Recommendation ITU-T P.501, Annex D [i.14]:

1) 48 kHz sampling frequency, low-pass filtered at 20 000 Hz (according to Recommendation ITU-T P.10/G.100 [i.15]) for FB or low-pass filtered at 14 000 Hz (according to Recommendation ITU-T G.191 [i.11]) for SWB (the British English and Dutch sample are only available in SWB):

   - P501_D_AM_fm_FB_48k.wav
   - P501_D_CN_fm_FB_48k.wav
   - P501_D_DU_fm_SWB_48k.wav
   - P501_D_EN_fm_SWB_48k.wav
   - P501_D_GE_fm_FB_48k.wav
   - P501_D_FI_fm_FB_48k.wav
   - P501_D_FR_fm_FB_48k.wav

- P501_D_IT_fm_FB_48k.wav

Usage:

- **To be used as reference signal P.863 FB (SWB).**
- Can be used as input signal in wideband and super-wideband measurement setups as in figure 3 of the present document.

2) 8 kHz sampling frequency, low-pass filtered at 3 800 Hz:

- P501_D_AM_fm_flat_08k.wav
- P501_D_CN_fm_flat_08k.wav
- P501_D_DU_fm_flat_08k.wav
- P501_D_EN_fm_flat_08k.wav
- P501_D_GE_fm_flat_08k.wav
- P501_D_FI_fm_flat_08k.wav
- P501_D_FR_fm_flat_08k.wav
- P501_D_IT_fm_flat_08k.wav

Usage:

- **To be used as reference signal P.863 NB.**

3) 8 kHz sampling frequency, Modified IRS send filtered according to Recommendation ITU-T P.830 [i.3]:

- P501_D_AM_fm_IRS_08k.wav
- P501_D_CN_fm_IRS_08k.wav
- P501_D_DU_fm_IRS_08k.wav
- P501_D_EN_fm_IRS_08k.wav
- P501_D_GE_fm_IRS_08k.wav
- P501_D_FI_fm_IRS_08k.wav
- P501_D_FR_fm_IRS_08k.wav
- P501_D_IT_fm_IRS_08k.wav

Usage:

- **NOT to be used as reference signal P.863 NB.**
- Can be used as input signal in narrowband measurement setups as in figure 3.

The samples listed above (with the sampling rates of 48 kHz and 8 kHz flat filtered and the 8 kHz sample IRSsend (mod) filtered) are all contained in Recommendation ITU-T P.501 [i.14], Annex D (https://www.itu.int/rec/T-REC-P.501-202005-I). It is recommended to use these samples for measurements without further processing.

History

Document history		
V1.1.1	October 2013	Publication
V1.2.1	November 2014	Publication
V1.3.1	March 2015	Publication
V1.4.1	September 2016	Publication
V1.5.1	August 2018	Publication
V1.6.1	March 2023	Publication