ETSI TR 103 138 V1.3.1 (2015-03)

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

Reference

RTR/STQ-00203m

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

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 (http://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 "**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

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

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.

Referenced documents which are not found to be publicly available in the expected location might be found at http://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.

Not applicable.

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
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
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 Narrowband and wideband 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 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 or in 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 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. Most of the recently standardized speech codecs are coding super-wideband signals. In practice, super-wideband can be seen as fullband for human speech, since there are no relevant signal parts above 14 000 Hz.

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. (P.862.2 [i.6]-wideband supports a dedicated wideband modus, however this measure was not established in the field and superseded by 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. However, testing in super-wideband mode gives the possibility to relate each limitation to an ideal sample (quasi fullband reference) and can 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 clip recorded in the studio the clips 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).

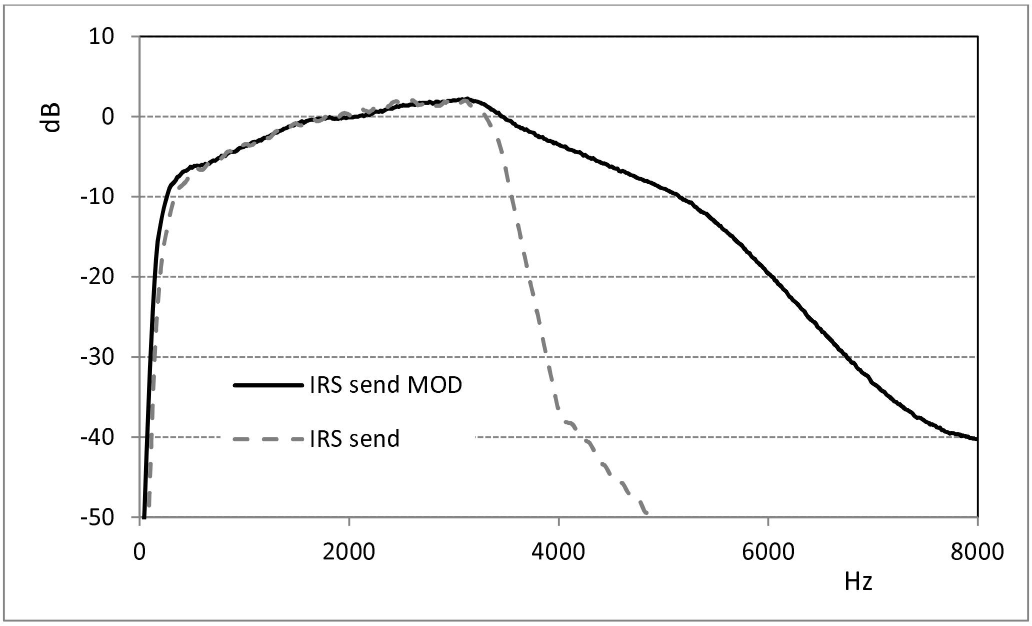

Figure 1: Frequency responses for IRS send and Modified IRS send filters

'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 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 G.191 [i.11].

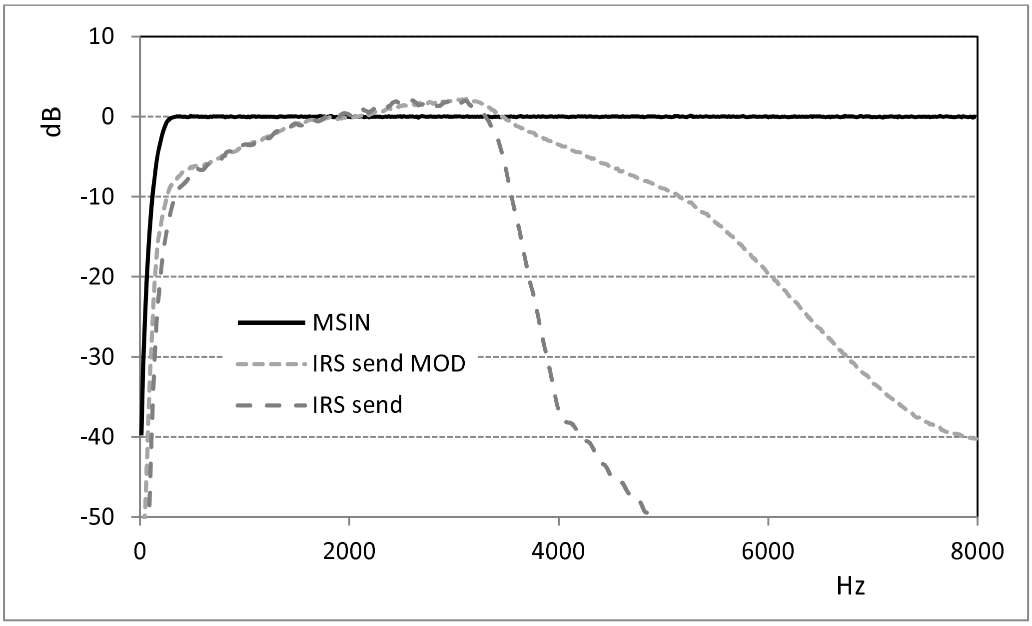

Figure 2: Frequency responses for IRS send and Modified IRS send filters compared to MSIN

5.2.2.3 Recommended filters to use in narrowband mobile test scenarios

In principle either filter can be used; however at the ISDN/PSTN interface a Modified IRS send prefiltering emulates the fixnet telephone device more accurately.

At the mobile side the situation is more difficult. Handset suppliers do not meet a standardized transmission characteristic as IRS send or MSIN. In the best case scenario the external microphone input at the headset connector has a flat characteristic.

In this case, a pre-filtered signal can be inserted (emulating the headset microphone). To avoid a mix of filter with the ISDN/PSTN side in mobile to PSTN tests, a Modified IRS send characteristic is recommended. This corresponds to point (2) in clause 4.3.2.

In case the microphone input at the mobile's headset connector does not have a flat characteristic, this characteristic will be added to the pre-filtering applied. Here are two possibilities:

1) The added characteristic is desired in order to take into account the handset's individual filter characteristic. The scores are predicting the quality with this particular phone under the assumption that an IRS send conformant headset is connected to this phone. The scores are phone dependent due to different filter characteristics applied by the phone used.

2) There is a compensation filter that equalizes the phone specific filter characteristic to a flat characteristic. Here the test connection emulates a phone that has a flat characteristic and is connected to an IRS send headset despite the actual individual characteristic of the phone. The sending path becomes more generic and the scores have less dependency on the actual filtering applied by a particular phone.

NOTE: Almost all of today's smart phones show already a flat sending characteristic (except a high pass at about 200 Hz).

5.2.3 Filter for wideband telephony test scenarios

5.2.3.1 14 kHz bandpass

As a filter for the so-called super-wideband speech, a bandpass from 50 Hz to 14 000 Hz with a flat bandpass is applied. A reference implementation is a filter described as '14KBP' in G.191 [i.11], that is the audio processing tool collection of ITU-T. A signal pre-filtered with this band-pass is recommended as input signal in a super-wideband test setup. Any limitation to this signal is considered as degradation as desired in a super-wideband scenario.

5.2.3.2 Recommendation ITU-T P.341

Recommendation ITU-T P.341 [i.12] describes test setups for telephone devices, there is also a tolerance scheme for common wideband filter defined. It is in principle a flat filter that filters wide band signals from 50 Hz to 7 000Hz. The tolerance scheme in 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 P.341 [i.12] filter has 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 P.341 [i.12] type filter upfront to the insertion of the speech signal is not recommended in case of insertion in 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.

5.2.3.3 Recommended filters to use in super-wideband mobile test scenarios

The use of a flat 50 Hz to 14 000 Hz filter is recommended. The pre-filtered speech is used as input signal. To cover the entire audio bandwidth, the signal has to be sampled at 48 kHz, 44,1 kHz or 32 kHz.

If there is an a-priori knowledge that the interface used for device reduces the bandwidth, a down-sampled signal can be used as input in this device. For example as input signal into a PSTN/ISDN interface card the flat super-wideband signal can be downsampled to 16 kHz or 8 kHz using a high quality down-sampling procedure. The same for a mobile device that is restricted to AMR-WB as highest audio bandwidth. Here the signal can be downsampled to 16 kHz. It still covers the bandwidth that is processed by this phone.

NOTE: The high quality downsampling routine in 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.

5.2.4 Reference signals

While the input signal might be pre-filtered and adapted to the access device, the reference signal as used for instrumental measures such as P.862 [i.4] or P.863 [i.8] remains largely unprocessed and has to preserve the minimum spectral range for the test case.

For tests with Recommendation ITU-T P.863 [i.8] in super-wideband mode (SWB), the reference signal has to be a flat signal that is just band-limited by the '14KBP' filter in G.191 [i.11] and may have a high-pass at 50 Hz to remove very low-frequency noises. It is the same signal as described in clause 5.1.2.3.

For tests in the NB mode Recommendation ITU-T P.863 [i.8] a references signal is allowed with either 48 kHz or with 8 kHz sampling frequency and a minimum bandwidth of up to 3 800 Hz. In practice, reference files for P.863 [i.8] NB measurements are sampled with 8 kHz which also provides a backwards compatibility to P.862.1 [i.5], where a reference signal sampled with 8 kHz is recommended too.

To receive such narrowband reference signals, the original signals have to be reduced in samplingrate and bandwidth. With the reconstruction filter as realized in 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.

5.3 Audio level

5.3.1 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 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. For speech at this nominal level all speech codecs in telephony are optimized and tested.

5.3.2 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 P.56 [i.13] 'Speech Voltmeter'. A speech clip at nominal level is normalized to -26 dB OVL according to P.56 [i.13].

5.3.3 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.

6 Scenarios

6.1 Narrowband-Measurement Land to Mobile

The speech sample is fed in on the fixed net side e.g. an ISDN device and transferred to the mobile. This is a narrowband scenario.

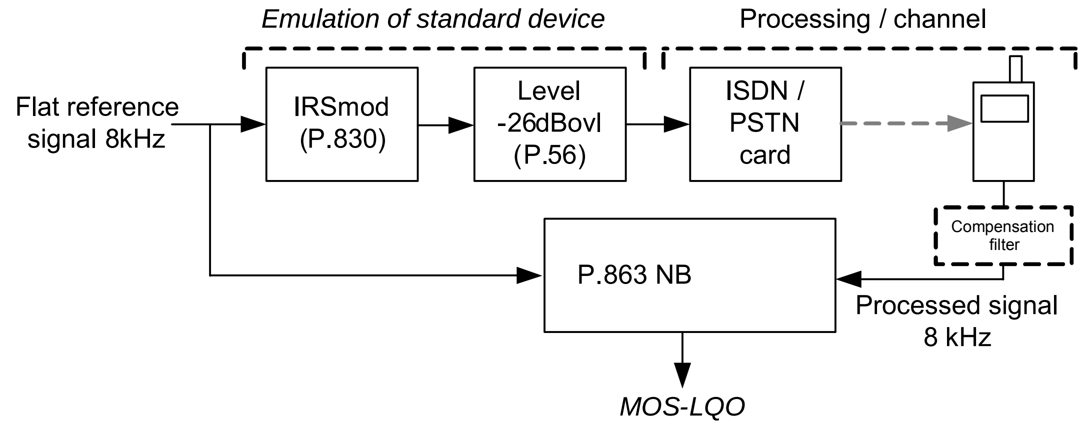

Figure 3

6.2 Narrowband-Measurement Mobile to Land

The clip is sent via the mobile to the receiver on the fixed net part. This is a narrowband scenario.

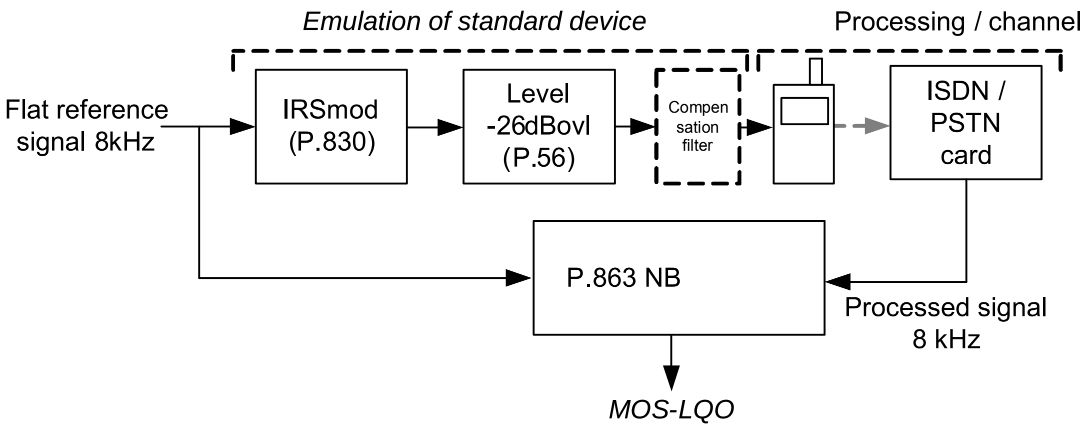

Figure 4

6.3 Mobile to Mobile

6.3.1 Narrowband

The clip is sent via the mobile to the receiver on the fixed net part. This is a narrowband scenario.

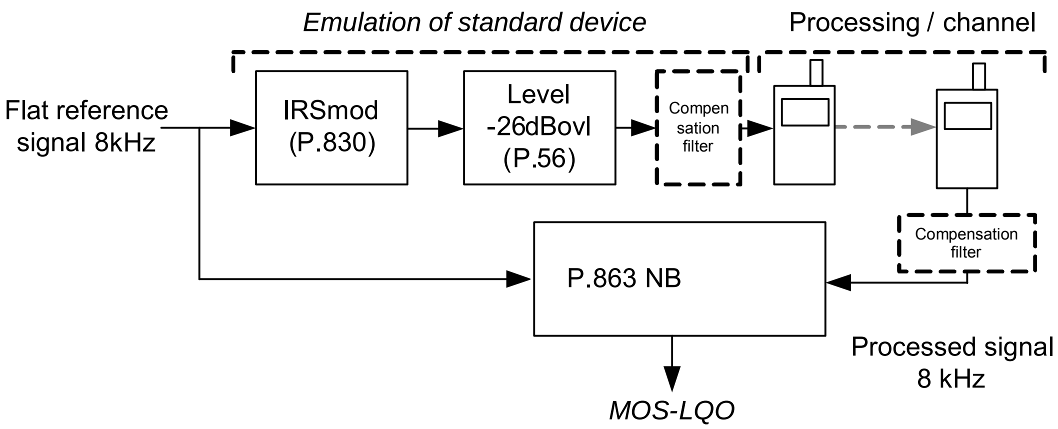

Figure 5

6.3.2 Wideband

The clip is sent via the mobile to the receiver also on the mobile. This is a wideband scenario.

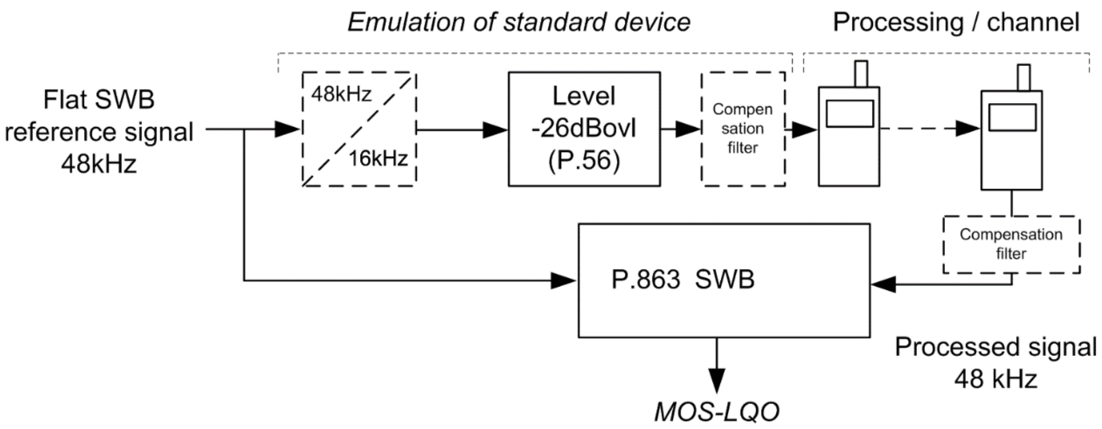

Figure 6

7 Synopsis

For a given test, either NB or WB, the following input and reference samples are to be used.

Table 1

SEND \ RECEIVE	ISDN	Mobile NB	Mobile WB
ISDN	Input: 8 kHz Pre-filter: Modified IRS send Ref: Flat NB 8 kHz P.863 [i.8]: NB	Input: 8 kHz Pre-filter: Modified IRS send Ref: Flat NB 8 kHz P.863 [i.8]: NB	Input: 8 kHz Pre-filter: Flat Ref: Flat SWB 48 kHz P.863 [i.8]: SWB
Mobile NB	Input: 8 kHz Pre-filter: Modified IRS send Ref: Flat NB 8 kHz P.863 [i.8]: NB	Input: 8 kHz Pre-filter: Modified IRS send Ref: Flat NB 8 kHz P.863 [i.8]: NB	Input: 8 kHz Pre-filter: Flat Ref: Flat SWB 48 kHz P.863 [i.8]: SWB
Mobile WB	Input: 16 / 48 kHz Pre-filter: Flat Ref: Flat SWB 48 kHz P.863 [i.8]: SWB	Input: 16 / 48 kHz Pre-filter: Flat Ref: Flat SWB 48 kHz P.863 [i.8]: SWB	Input: 16 / 48 kHz Pre-filter: Flat Ref: Flat SWB 48 kHz P.863 [i.8]: SWB

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_103138v010301p0.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:

- Dutch DU_P501_fm
- British English EN_P501_fm
- German GE_P501_fm
- Finnish FI_P501_fm
- French FR_P501_fm
- Italian IT_P501_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 6s 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 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 P.501 [i.14], where the boundaries are shown as grey dashed lines.

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

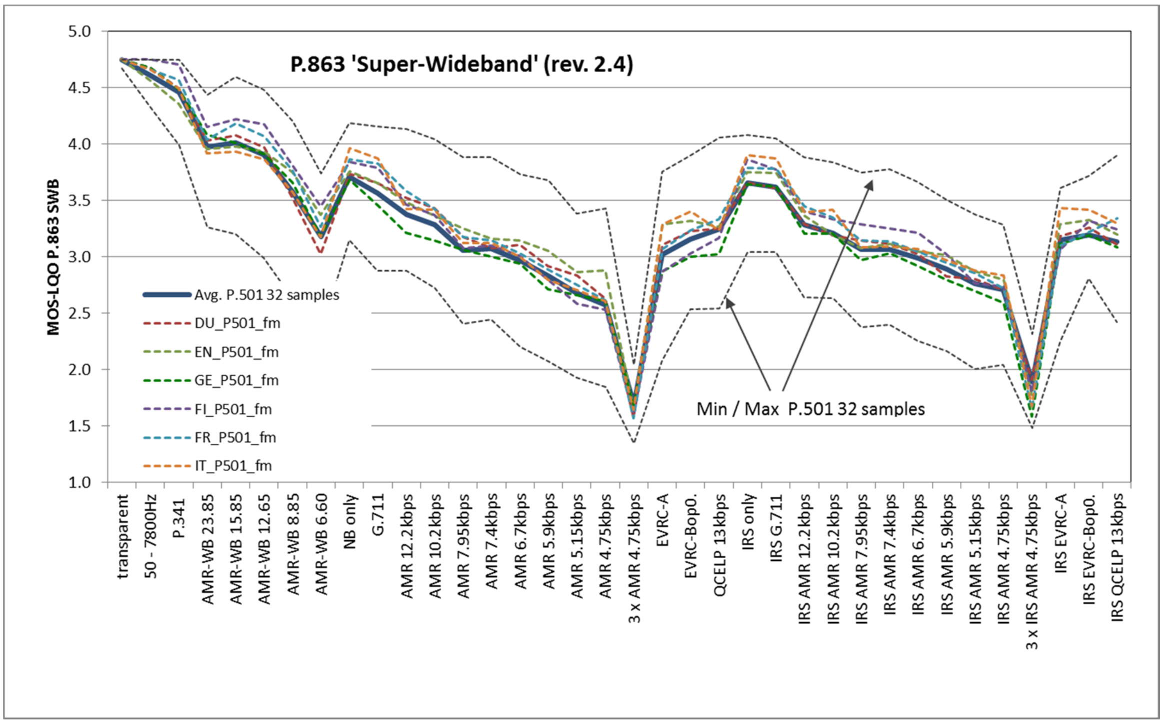

Figure C.1: P.863 [i.8] SWB example results of the composed speech samples in comparison to the average of all 32 samples from P.501 [i.14] for specific codec processing conditions

A corresponding analysis for P.863 [i.8] in narrowband mode shows a very good correspondence too. The processing is in line with Figures 3 to 5 of the present document.

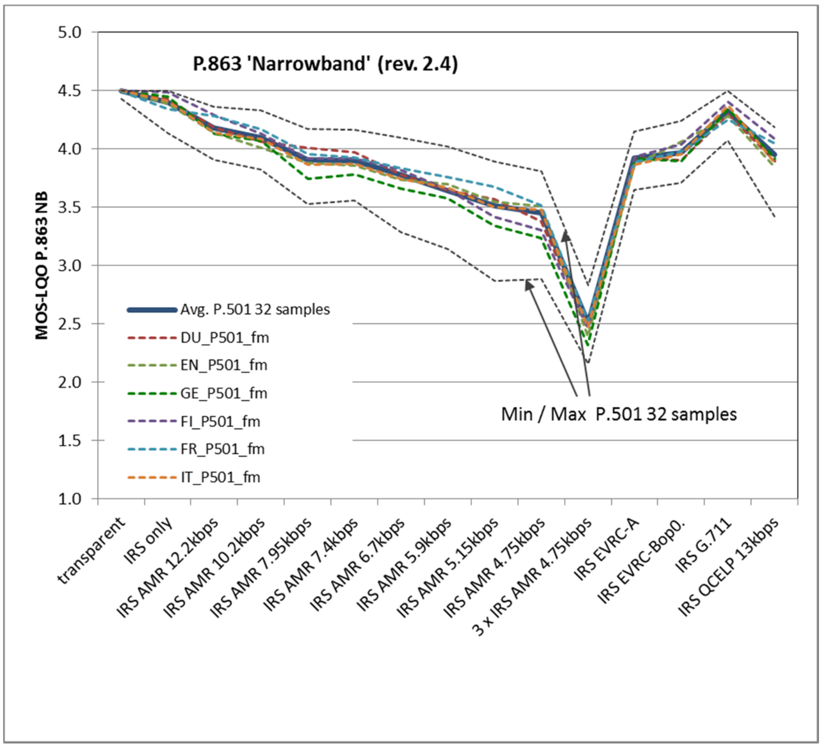

Figure C.2: P.863 [i.8] NB example results of the composed speech samples in comparison to the average of all 32 samples from P.501 [i.14] for specific codec processing conditions

C.4 Technical specification

To meet the recommendations as given in clauses 6 and 7 of the present document, all samples are provided in different sampling rates and pre-filters. As interpolation low-pass for re-sampling the coefficients as provided in Annex A are used.

1) 48 kHz sampling frequency, low-pass filtered at 14 000 Hz (according to Recommendation ITU-T G.191 [i.11]) as in SWB specification:
   - `DU_P501_fm_flat_48k.wav`
   - `EN_P501_fm_flat_48k.wav`
   - `GE_P501_fm_flat_48k.wav`
   - `FI_P501_fm_flat_48k.wav`
   - `FR_P501_fm_flat_48k.wav`
   - `IT_P501_fm_flat_48k.wav`

   Usage:

   - **To be used as reference signal P.863 SWB**
   - Can be used as input signal in wideband measurements setups as in Figure 6 of the present document.

2) 16 kHz sampling frequency, low-pass filtered at 7 600 Hz
   - `DU_P501_fm_flat_16k.wav`
   - `EN_P501_fm_flat_16k.wav`
   - `GE_P501_fm_flat_16k.wav`
   - `FI_P501_fm_flat_16k.wav`
   - `FR_P501_fm_flat_16k.wav`
   - `IT_P501_fm_flat_16k.wav`

   Usage:

   - **NOT to be used as reference signal P.863 SWB**
   - Can be used as input signal in wideband measurements setups as in Figure 6 of the present document.

3) 8 kHz sampling frequency, low-pass filtered at 3 800 Hz
   - `DU_P501_fm_flat_08k.wav`
   - `EN_P501_fm_flat_08k.wav`
   - `GE_P501_fm_flat_08k.wav`
   - `FI_P501_fm_flat_08k.wav`
   - `FR_P501_fm_flat_08k.wav`
   - `IT_P501_fm_flat_08k.wav`

   Usage:

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

4) 8 kHz sampling frequency, Modified IRS send filtered according to Recommendation ITU-T P.830 [i.3]
   - `DU_P501_fm_IRSm_08k.wav`
   - `EN_P501_fm_IRSm_08k.wav`
   - `GE_P501_fm_IRSm_08k.wav`
   - `FI_P501_fm_IRSm_08k.wav`
   - `FR_P501_fm_IRSm_08k.wav`
   - `IT_P501_fm_IRSm_08k.wav`

   Usage:

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

The samples listed above (with the sampling rates of 48, 16 and 8 kHz flat filtered and the 8 kHz sample IRSsend (mod) filtered) are all contained in the archive tr_103138v010301p0.zip which accompanies the present document. 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