ETSI TS 101 512 V8.0.0 (2000-08)

Technical Specification

Digital cellular telecommunications system (Phase 2+); Minimum Performance Requirements for Noise Suppresser Application to the AMR Speech Encoder (GSM 06.77 version 8.0.0 Release 1999)

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

Reference

DTS/SMG-110677Q7

Keywords

Global System for Mobile communications (GSM), speech, Adaptive Multi rate (AMR), Noise suppression

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://www.etsi.org/ipr).

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 Specification (TS) has been produced by the Special Mobile Group (SMG).

The present document specifies minimum performance requirements for Noise Suppression for the Adaptive Multi Rate (AMR) codec within the digital cellular telecommunications system.

The contents of the present document is subject to continuing work within SMG and may change following formal SMG approval. Should SMG modify the contents of the present document, it will be republished by ETSI with an identifying change of release date and an increase in version number as follows:

Version 8.x.y

where:

8 GSM Phase 2+ Release 1999.

x the second digit is incremented for changes of substance, i.e. technical enhancements, corrections, updates, etc.;

y the third digit is incremented when editorial only changes have been incorporated in the specification.

1 Scope

The present document specifies recommended minimum performance requirements for noise suppression algorithms intended for application in conjunction with the AMR speech encoder. This specification is for guidance purposes. Noise Suppression is intended to enhance the speech signal corrupted by acoustic noise at the input to the AMR speech encoder.

The use of this recommended minimum performance requirements specification is not mandatory except for those solutions intended to be endorsed by SMG11.

It is the intention of SMG11 to perform analysis and validation of any AMR noise suppression solution which is voluntarily brought to the attention of SMG11 in the future, using the requirements set out in this specification to facilitate such an analysis. In order for SMG11 to endorse such a solution, SMG11 must confirm that all the recommended minimum performance requirements are met.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.
- For this Release 1999 document, references to GSM documents are for Release 1999 versions (version 8.x.y).

[1] CCITT Recommendations I.130 (1988): "General modelling methods - Method for the characterisation of telecommunications services supported by an ISDN and network capabilities of an ISDNî".

[2] GSM 01.04 (ETR 350): "Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms".

3 Definitions and abbreviations

GSM 01.04 (ETR 350) [2] provides a list of abbreviations and acronyms used in GSM specifications. For the purposes of the present document, the following definitions and abbreviations also apply:

3.1 Definitions

None

3.2 Abbreviations

AMR Adaptive Multi-Rate
AMR/NS Combination of the AMR speech codec and the Noise Suppression function
NS Noise Suppression

4 Description of Noise Suppression applied to AMR

Noise Suppression for the AMR codec is a feature designed to enhance speech quality in a range of environments where there is significant (acoustic) background noise. The noise suppression function is a pre-processing module that is used to improve the signal to noise ratio of a speech signal prior to voice coding. In so doing it may use functions and/or data from the AMR speech encoding function. This specification defines recommended minimum performance requirements for such a function when it is implemented in the mobile station (operating on the uplink speech signal).

The AMR Speech decoder should not be altered by the Noise Suppression function.

It shall be possible to disable the operation of the noise suppression algorithm using signalling when commanded by the network.

4.1 Applicability of Noise Suppression to Basic Services.

This feature shall be applicable (as an option) to all speech calls where the narrowband AMR codec is utilised. Provision of the feature in AMR-capable mobile stations is a manufacturer dependent option. The network shall be able to enable or disable this noise suppression function both at call set-up and in call. [Signalling between network and mobile to allow this control is under study in SMG2 WPA].

5 Requirements to be assessed by Objective Means

5.1 Bit Exactness of the Speech Encoder

The Noise Suppression shall be implemented as a separate pre-processing module prior to the speech encoding. The functionality and all internal states, tables and variables of the speech encoder shall remain unaltered by the Noise Suppression function.

The Noise Suppression should be implemented as a stand-alone pre-processing module operating on the 160 samples input speech buffer to the speech encoder according to Figure 1.

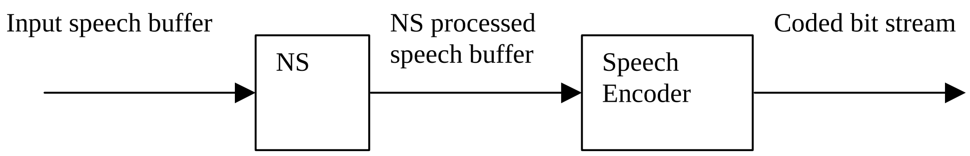

Figure 1: Noise Suppression implementation

Alternatively, for implementation in conjunction with the bit-exact fixed point C reference code [GSM 06.73] the NS module may operate on the pre-processed input speech buffer “old_speech[L_TOTAL]” in the structure “cod_amrState” in the AMR C code [GSM 06.73] after the pre-processing module (sample down-scaling and input high pass filtering) of the speech encoder. The bit-integrity of the speech encoder for this implementation shall be verified according to Figure 2 where the signals at Test Points 1 and 2 shall be identical for any input signal and the Reference Encoder is the part of [GSM 06.73] after the pre-processing module. Note: implementation in conjunction with the AMR floating point C code is for further study.

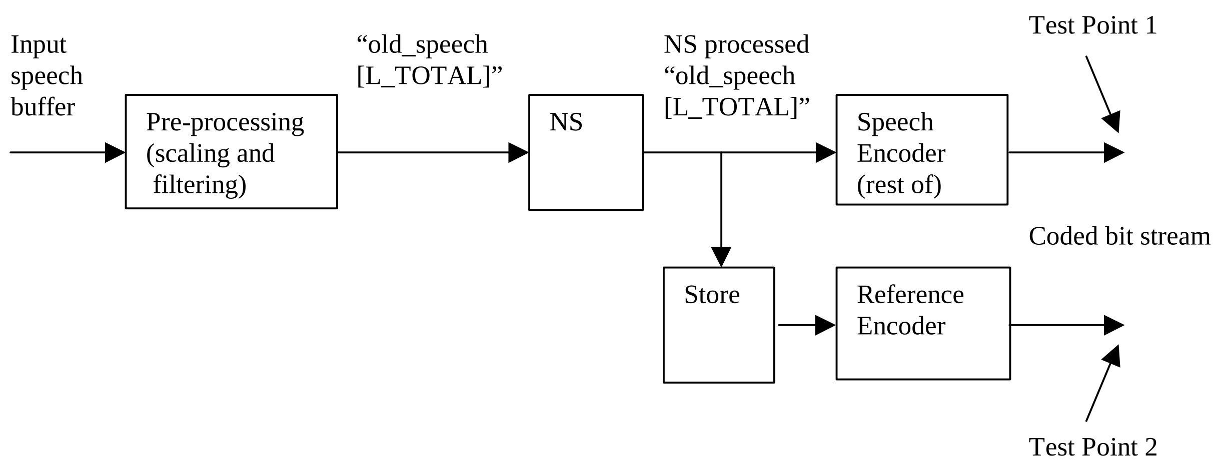

Figure 2: Verification of AMR speech encoder bit-exactness for embedded NS implementations

5.2 Bit Exactness of the Speech Decoder

The AMR speech decoder shall remain unaltered by the Noise Suppression function.

5.3 Impact on Speech Path Delay

The one way algorithmic delay due to the activation of AMR noise suppression shall be no more than 5ms in excess of the delay inserted by the AMR speech codec. In the handsfree case, this delay is part of the 39ms delay specified in GSM 03.50.

The total additional delay (comprising of algorithmic and processing delays) shall not exceed 10ms. The processing delay is calculated using the following formula with E*S*P set to 50.

$$\text{delay(proc)} = \text{WMOPS} * 20/(E * S * P)$$

where WMOPS = complexity in weighted operations per second evaluated through the theoretical worst case. (Direct means of measurement of total delay is for further study.).

5.4 Impact on Channel Activity

The AMR speech codec with noise suppression activated should not significantly increase channel activity when used in conjunction with DTX.

Channel activity increase will be measured thanks to the Voice Activity factor (VAF), defined as follows.

Let x be the VAF measured by the AMR VAD as an averaged value on all clean speech signals

Let y be the VAF measured by the AMR VAD without AMR NS active as an averaged value on all clean speech + noise signals (where the applicable clean speech signal is the speech signal used in the measure of x).

Let w be the VAF measured by the AMR VAD with AMR NS active as an averaged value on all clean speech +noise signals (where the applicable clean speech signal is the speech signal used in the measure of x). w is required to be not significantly more than the maximum of y and x. Any case where w is greater than y should be further investigated.

These requirements shall apply to all standardized AMR VADs. (w,x,y) are determined using all VADs, and the requirements are checked relatively to each AMR VAD independently.

The definition of upper limits on VAF increase and attendant confidence intervals are for further study.

6 Requirements to be assessed by subjective tests

6.1 Impact on Speech Quality

The following performance requirements are stated under the assumption that the noise suppresser is tested as an integral part of the AMR speech codec with the speech codec operating at the rates defined within the test plan ([reference to be added when test plan is available]) .The performance requirements must be met for all these stated speech codec rates.

6.1.1 Initial Convergence Time

The initial convergence time shall be a maximum of T seconds with T equal to 2s. The definition of this time interval shall be understood strictly in accordance with its means of use in subjective listening experiments. Its use shall be defined by a process whereby the first T seconds of each sample processed through the AMR speech codec with and without noise suppression active, is deleted before presentation to listeners. It is assumed that this process does not reduce intelligibility, or introduce clipping or similar effects into the resultant speech plus noise material.

6.1.2 No Degradation in Clean Speech

The noise suppression function must not have a statistically significant distorting effect on clean speech, in comparison with the performance of the AMR codec without noise suppression applied. This requirement also applies when VAD/DTX is active.

The requirement is checked with the use of a paired comparison test where the requirement is met if AMR/NS is preferred or equal to AMR within the 95 % confidence interval.

6.1.3 No degradation of Speech and no Undesirable Effects in Residual Noise in Conditions with Background Noise (*residual noise = background noise after AMR/NS)*

The noise suppression function must not introduce any degradation of speech and no undesirable effects in the residual noise, when there is (acoustic) background noise in the speech signal. This requirement also applies when VAD/DTX is active.

The requirement is checked with the use of a modified ACR test with specific instructions where the requirement is met if AMR/NS is better than or equal to AMR within the 95 % confidence interval in all conditions.

6.1.4 Quality Impact compared to AMR

The AMR speech codec with noise suppression activated must produce an output in noisy speech which is preferred amongst test listeners with statistical significance, compared to the case where noise suppression is not used. This requirement also applies when VAD/DTX is active.

The requirement is checked with the use of a CCR test where the requirement is met if AMR/NS is preferred to AMR within the 95 % confidence interval in at least 4 of the 6 (*number of test conditions to be confirmed)* conditions tested. Preference or equality within the 95 % confidence interval is required for the remaining conditions.

[Following requirements for SNR improvement are to be confirmed.]

Additionally, it is required that the subjective SNR improvement as measured by the methodology [Ref 1] where the measure is conducted on all CCR tests [Ref. 2] meets the following requirements:

(a) In at least 2 of the 6 conditions tested the SNR improvement shall not be less than 6dB within the 95% confidence interval.

(b) In at least 2 of the remaining 4 conditions the SNR improvement shall not beless than 4dB within the 95% confidence interval.

[NOTE: Refs 1 and 2 to be added; Ref 1 references the SNR improvement measurement methodology, Ref. 2 references the test plan, currently under development, designed to test the requirements in this specification.]

7 Performance Objectives assessed by Objective Measures

7.1 Impact on Active Speech Level

The AMR speech codec with noise suppression activated must not significantly alter the active speech level.

The requirement is checked with the use of a P.56 speech level meter (the use of which remains for further study). Let x be the averaged level of the clean speech material for one experiment and let y be the averaged level of the processed material with AMR NS activated for the same experiment. The requirement is met if the absolute difference between x and y is less than [2] dB for all experiments. *The processed material should not be normalised to the nominal speech level before the listening tests.*

Note that this requirement does not preclude the use of active level control.

7.2 Objective Speech Quality Measures

The objective measures of noise power level reduction (NPLR) and signal-to-noise ratio improvement (SNRI) defined in Annex 1 are to be used to characterise the performance of the AMR/NS solution. Objectives are defined for these measures in the following table. These measures will be used to provide additional information only and are not to be considered to be requirements.

C source code is attached to this specification which shall be used to undertake these measurements.

Objective quality measure/test condition	**Performance objective**
NPLR *Assessment*: To be evaluated using a predefined set of material (as used in the AMR/NS Selection Phase) comprising speech mixed with stationary car noise in the SNR conditions of 6 dB and 15 dB, following otherwise the guidelines set forth in [Annex 1].	–7 dB or lower
SNRI *Assessment:* To be evaluated using a predefined set of material (as used in the AMR/NS Selection Phase) comprising speech mixed with stationary car noise in the SNR conditions of 6 dB and 15 dB, following otherwise the guidelines set forth in [Annex 1].	6 dB or higher

8 Interaction with supplementary services

8.1 General

This clause defines requirements regarding the interactions between GSM supplementary services and the Noise Suppression Feature.

The application of Noise Suppression shall not interfere with the provision or invocation of any supplementary services.

8.2 Explicit Call Transfer (ECT)

No adverse interaction. If the new party is a mobile station with support for the Noise Suppression feature, the noise suppression feature shall be invoked.

8.3 Call wait/Call hold.

No interaction.

8.4 Multiparty

No interaction.

8.5 Service Announcements

No interaction.

9 Interaction with Alternate and Followed by services

There shall be no impact on data transmission due the Noise Suppression Feature

10 Interaction with other speech services

There is no requirement for Noise Suppression in ASCI services.

11 Interaction with DTMF and other signalling tones

DTMF and other signalling tones transmission performance during the application of Noise Suppression shall be no worse than the case where Noise Suppression is turned off.

12 Interaction with Lawful Intercept

In the case where lawful intercept is required in a call where Noise Suppression is activated, the Noise Suppression shall not cause any degradation in the speech quality received by the A and B parties.

13 Interaction with TFO

No interaction.

Annex A (informative): Method for generating Objective Performance Measures

This annex presents an objective methodology for characterising the performance of noise suppression (NS) methods. Two objective measures are presented to be used for characterising NS solutions complying with the AMR/NS specification.

A.1 Objective measures and test signals

A1.1 Notations

The following notations are used in this document:

- The operator AMR(·) corresponds to applying the AMR speech encoder and decoder on the input.
- The operator NR(·) corresponds to applying the NS algorithm, and the AMR speech encoder and decoder on the input.
- The clean speech signals are referred to as $\mathbf{s_I}$, $i = 1$ *to* I.
- The noise signals are referred to as $\mathbf{n_j}$, $j = 1$ *to* J.
- The noisy speech test signals are referred to as $\mathbf{d_{Ij} = \beta_{ij}(SNR)\, n_j + s_i}$, $i = 1$ *to* I, $j = 1$ *to* J, where d_{ij} is built by adding s_i and n_j with a pre-specified SNR as presented below.
- The processed signal are referred to as $\mathbf{y_{ij} = NR\,(d_{ij})}$.
- The reference signal in the calculations shall be either the noisy speech test signal d_{ij} itself or d_{ij} processed by the AMR speech codec without NS processing. The latter signal will be referred to as $\mathbf{c_{Ij} = AMR\,(d_{ij})}$, $i = 1$ *to* I, $j = 1$ *to* J. The relevant reference signal will be indicated in the formulation of each objective measure below.
- The notation *Log*(·) indicates the decimal logarithm.
- $\mathbf{\beta_{ij}(SNR)}$ is the scaling factor to be applied to the background noise signal $\mathbf{n_i}$ in order to have a ratio **SNR** (in dB) between the clean speech signal $\mathbf{s_i}$ and $\mathbf{n_j}$. The scaling of the input speech and noise signals is to be carried according to the following procedure:
  - The clean speech material is scaled to a desired dBov level with the ITU-T recommendation P.56 speech voltmeter, one file at a time, each file including a sequence of one to four utterances from one speaker.
  - A silence period of 2 s is inserted in the beginning of each of the resulting files to make up augmented clean speech files.
  - Within each noise type and level, a noise sequence is selected for every speech utterance file, each with the same length as the corresponding speech files, and each noise sequence is stored in a separate file.
  - Each of the noise sequences is scaled to a dBov level leading to the SNR condition corresponding to the $\mathbf{\beta_{ij}(SNR)}$ value in each of the test cases by applying the RMS level based scaling according to the P.56 recommendation.
- The determination of which frames contain active speech is to be carried out with reference to the ITU-T recommendation P.56 active speech level measurement and is related to the classification of the frames into the presented speech power classes which is explained below.

A1.2 Test material

The test material should manifest at least the following extent:

- Clean speech utterance sequences: 6 utterances from 4 speakers - 2 male and 2 female - totalling 24 utterances
- Noise sequences:
  - car interior noise, 120 km/h, fairly constant power level;
  - street noise, slowly varying power level.

Special care should be taken to ensure that the original samples fulfill the following requirements:

- the clean speech signals are of a relatively constant average (within sample, where 'sample' refers to a file containing one or more utterances) power level;
- the noise signals are of a short-time stationary nature with no rapid changes in the power level and no speech-like components.

The test signals should cover the following background noise and SNR conditions:

- car noise at 3 dB, 6 dB, 9 dB, 12 dB and 15 dB;
- street noise at 6 dB, 9 dB, 12 dB, 15 dB and 18 dB.

A feasible subset of these conditions giving a practically useful indication of the achieved performance would be:

- car noise at 6 dB and 15 dB;
- street noise at 9 dB and 18 dB.

The samples should be digitally filtered before NS and speech coding processing by the MSIN filter to become representative of a real cellular system frequency response.

A1.3 Proposal for objective measures for NS performance assessment

Assessment of SNR improvement level: The SNR improvement measure, ***SNRI***, measures the SNR improvement achieved by the NS algorithm. SNR improvement is calculated separately in three frame power gated factors of active speech signal, namely, high, medium and low power constituents of the signal. These categories are used to characterise the effect of the NS processing on speech, allowing to distinguish the effect on strong, medium and weak speech. In addition to calculating the SNR improvement separately on the three categories, they are used to form an aggregate measure.

The calculation is here presented for the high power speech class:

For each background noise condition j

For each speaker i

Construct a noisy input signal d_{ij} as follows:

$$d_{ij}(n) = \beta_{ij}\ n_j(n) + s_i(n)$$

where β_{ij} depends on the SNR condition according to the procedure described above

c_{ij} = AMR (d_{ij})

y_{ij} = NR (d_{ij})

$$\mathrm{SNRout\_h_{ij}} = \frac{\xi + \frac{1}{K_{sph}} \sum_{k=k_{sph,1}}^{k_{sph,K_{sph}}} \sum_{n=k\cdot 80}^{k\cdot 80+79} \mathrm{y_{ij}^2(n)}}{\xi + \frac{1}{K_{nse}} \sum_{l=k_{nse,1}}^{k_{nse,K_{nse}}} \sum_{n=l\cdot 80}^{l\cdot 80+79} \mathrm{y_{ij}^2(n)}} - 1$$

$$\mathrm{SNRin\_h_{ij}} = \frac{\xi + \frac{1}{K_{sph}} \sum_{k=k_{sph,1}}^{k_{sph,K_{sph}}} \sum_{n=k\cdot 80}^{k\cdot 80+79} \mathrm{c_{ij}^2(n)}}{\xi + \frac{1}{K_{nse}} \sum_{l=k_{nse,1}}^{k_{nse,K_{nse}}} \sum_{n=l\cdot 80}^{l\cdot 80+79} \mathrm{c_{ij}^2(n)}} - 1$$

$$\mathrm{SNRI\_h_{ij}} = \begin{cases} 0 & ; \quad \mathrm{SNRout\_h_{ij}} \le \xi \vee \mathrm{SNRin\_h_{ij}} \le \xi \\ 10 \cdot \left[\mathrm{Log}\left(\mathrm{SNRout_{ij}}\right) - \mathrm{Log}\left(\mathrm{SNRin_{ij}}\right)\right] & ; \quad \mathrm{else} \end{cases} \tag{1}$$

where k_{sph} and K_{sph} are the index and the total number of frames containing speech of a high power

k_{nse} and K_{nse} are the corresponding index and total number of noise only frames

$\xi > 0$ is a constant that should be set at 10^{-5}

$\mathrm{SNRI\_m_{ij}}$ correspondingly for medium power frames

$\mathrm{SNRI\_l_{ij}}$ correspondingly for low power frames

$$\mathrm{SNRI_{ij}} = \frac{1}{K_{sph} + K_{spm} + K_{spl}} \left(K_{sph} \cdot \mathrm{SNRI\_h_{ij}} + K_{spm} \mathrm{SNRI\_m_{ij}} + K_{spl} \mathrm{SNRI\_l_{ij}}\right) \tag{2}$$

$$\mathrm{SNRI_j} = \frac{1}{\mathrm{I}} \sum_{i=1}^{I} \mathrm{SNRI_{ij}} \tag{3}$$

$$\mathrm{SNRI} = \frac{1}{\mathrm{J}} \sum_{j=1}^{J} \mathrm{SNRI_j} \tag{4}$$

In addition, measures for the SNR improvement in the high, medium and low power speech classes (SNRI_h, SNRI_m, SNRI_l, respectively) shall be recorded based on the following formulae:

$$\mathrm{SNRI\_h} = \frac{1}{J} \sum_{j=1}^{J} \mathrm{SNRI\_h_j} = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\mathrm{I}} \sum_{i=1}^{I} \mathrm{SNRI\_h_{ij}} \tag{5}$$

$$\mathrm{SNRI\_m} = \frac{1}{J} \sum_{j=1}^{J} \mathrm{SNRI\_m_j} = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\mathrm{I}} \sum_{i=1}^{I} \mathrm{SNRI\_m_{ij}} \tag{6}$$

$$\mathrm{SNRI\_l} = \frac{1}{J} \sum_{j=1}^{J} \mathrm{SNRI\_l_j} = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\mathrm{I}} \sum_{i=1}^{I} \mathrm{SNRI\_l_{ij}} \tag{7}$$

It is, in addition, informative to record separately the noise type specific SNR improvement measures, namely, $\mathrm{SNRI\_h_j}$, $\mathrm{SNRI\_l_j}$, $\mathrm{SNRI\_m_j}$ and $\mathrm{SNRI_j}$ for each j.

To determine which frames belong to high, medium and low power classes of active speech and which present pauses in the speech activity (noise only), the active speech level (in dB) sp_lvl of the noise free speech $s_i(n)$ is first determined according to the ITU-T recommendation P.56. Thereafter, the frames are classified into the four classes as follows:

for all signal frames k:

$$\text{sp\_pow}(k) = 10\log\left[\max\left(\varepsilon, \frac{\sum_{n=k\cdot 80}^{k\cdot 80+79}(s_i(n))^2}{80}\right)\right] \tag{8}$$

if $\text{sp\_pow}(k) \geq \text{sp\_lvl} + \text{th\_h}$

$$\left\{k_{sph,\text{length}(\text{k}_{\text{sph}})+1}\right\} = \left\{k_{sph,\text{length}(\text{k}_{\text{sph}})}, k\right\}$$

else if $\text{sp\_pow}(k) \geq \text{sp\_lvl} + \text{th\_m}$

$$\left\{k_{spm,\text{length}(\text{k}_{\text{sp}\,m})+1}\right\} = \left\{k_{spm,\text{length}(\text{k}_{\text{sp}\,m})}, k\right\} \tag{9}$$

else if $\text{sp\_pow}(k) \geq \text{sp\_lvl} + \text{th\_l}$

$$\left\{k_{spl,\text{length}(\text{k}_{\text{sp}\,l})+1}\right\} = \left\{k_{spl,\text{length}(\text{k}_{\text{sp}\,l})}, k\right\}$$

else if $\text{sp\_lvl} + \text{th\_nl} \leq \text{sp\_pow}(k) < \text{sp\_lvl} + \text{th\_nh}$

$$\left\{k_{nse,\text{length}(\text{k}_{nse})+1}\right\} = \left\{k_{nse,\text{length}(\text{k}_{nse})}, k\right\}$$

where $\varepsilon > 0$ is a constant whose value shall be such that in the dB scale, it shall be below sp_lvl + th_nl; a value of 10^{-7} should be used if sp_lvl = -26 dBov and th_nl = -34 dB, as proposed below:

th_h, th_m, th_l are pre-determined lower threshold power levels for classifying the speech frames to the high, medium, and low power classes, correspondingly.

The following notes on the formulation of the frame classification are made:

- The lower bound for the power of the noise-only class of frames is motivated by a desire to restrict the analysis to noise frames that are among or close the speech activity, hence excluding long pauses from the analysis. This makes the analysis concentrate increasingly on the effects encountered during speech activity.
- In poor SNR conditions, the noise power level may occur to be higher than the lower bound of some of the speech power classes. However, even in this case, the information of the effect on the low power portions of speech may be informative. Another way of formulating the measure might be to make the power thresholds dependent on the noise level. This would, however, restrict the comparability of the SNR improvement figures of the different classes over experiments with different background noise content.
- The presented method of classifying the speech frames in the designated classes and, hence, determining values for the SNR improvement measures, is only applicable if all the used power level threshold values are higher than the corresponding power threshold level derived in the speech level measurement referred to above.

The scaling for the clean speech material should be determined optimally so that the dynamics of the 16 bit arithmetic system is efficiently used but no waveform clipping is produced. Typically, a normalisation to the active speech level of –26 dBov is preferable. In such a case, the following values should be used for the power class thresholds:

th_h = -1 dB

th_m = -10 dB

th_l = -16 dB (10)

th_nh = -19 dB

th_nl = -34 dB

Assessment of noise power level reduction. The noise power level reduction ***NPLR*** measure relates to the capability of the NS method to attenuate the background noise level.

The ***NPLR*** measure is calculated as follows:

For each background noise condition j

For each speaker i

Construct a noisy input signal d_{ij} as follows:

$d_{ij}(n) = \beta_{ij}\ n_j(n) + s_i(n)$

where β_{ij} depends on the SNR condition according to the procedure described above

c_{ij} = AMR (d_{ij})

y_{ij} = NR (d_{ij})

$$NPLR_{ij} = 10 \cdot \left\{ \mathrm{Log}\left[\xi + \frac{1}{K_{nse}} \sum_{k=k_{nse,1}}^{k_{nse,K_{nse}}} \sum_{n=k\cdot 80}^{k\cdot 80+79} \mathrm{y}_{\mathrm{ij}}^2(\mathrm{n}) \right] - \mathrm{Log}\left[\xi + \frac{1}{K_{nse}} \sum_{l=k_{nse,1}}^{k_{nse,K_{nse}}} \sum_{n=l\cdot 80}^{l\cdot 80+79} \mathrm{c}_{\mathrm{ij}}^2(\mathrm{n}) \right] \right\}, \qquad (11)$$

where $\xi > 0$ is a constant that should be set at 10^{-5};

k_{nse} and K_{nse} are the corresponding index and total number of noise only frames

$$NPLR_j = \frac{1}{\mathrm{I}} \sum_{i=1}^{I} NPLR_{ij} \qquad (12)$$

$$NPLR = \frac{1}{\mathrm{J}} \sum_{j=1}^{J} NPLR_j \qquad (13)$$

Furthermore, it is informative to record separately the noise type specific NPLR measures, or $NPLR_j$, for each j.

Comparison of *SNRI* and *NPLR*. A comparison of the ***SNRI*** and ***NPLR*** measures can be used to acquire an indication of possible speech distortion produced by the tested NS method. If the ***NPLR*** parameter assumes clearly higher values than ***SNRI***, it can be expected that the NS candidate causes distortion to speech. This relation, however, should always be verified through a comparison with subjective test results.

Annex B (informative): Change Request History

SMG#	Tdoc SMG	Spec	CR	Cat	PH	Vers	New Vers	Subject

History

Document history		
V8.0.0	July 2000	Publication