ETSI EG 202 396-3 V1.7.1 (2018-03)

Speech and multimedia Transmission Quality (STQ);
Speech Quality performance
in the presence of background noise;
Part 3: Background noise transmission -
Objective test methods

Reference
REG/STQ-270

Keywords
noise, QoS, quality, speech

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

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

The present document is a deliverable of ETSI Specialized Task Force (STF) 294 entitled: "Improving the quality of eEurope wideband speech applications by developing a performance testing and evaluation methodology for background noise transmission".

The present document is part 3 of a multi-part deliverable covering Speech and multimedia Transmission Quality (STQ); Speech Quality performance in the presence of background noise, as identified below:

Part 1: "Background noise simulation technique and background noise database";

Part 2: "Background noise transmission - Network simulation - Subjective test database and results";

Part 3: "Background noise transmission - Objective test methods".

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.

1 Scope

The present document aims to identify and define testing methodologies which can be used to objectively evaluate the performance of narrowband and wideband terminals and systems for speech communication in the presence of background noise.

Background noise is a problem in mostly all situations and conditions and need to be taken into account in both, terminals and networks. The present document provides information about the testing methods applicable to objectively evaluate the speech quality in the presence of background noise. The present document includes:

- The description of the experts post evaluation process chosen to select the subjective test data being within the scope of the objective methods.
- The results of the performance evaluation of the currently existing methods described in Recommendations ITU-T P.862 [i.16] and P.862.1 [i.17] and in TOSQA2001 [i.19] which is chosen for the evaluation of terminals in the framework of ETSI VoIP speech quality test events [i.8], [i.9], [i.10] and [i.11].
- The method which is applicable to objectively determine the different parameters influencing the speech quality in the presence of background noise taking into account:
  - the speech quality;
  - the background noise transmission quality;
  - the overall quality.
- The present document is to be used in conjunction with:
  - ETSI ES 202 396-1 [i.1] which describes a recording and reproduction setup for realistic simulation of background noise scenarios in lab-type environments for the performance evaluation of terminals and communication systems.
  - ETSI EG 202 396-2 [i.2] which describes the simulation of network impairments and how to simulate realistic transmission network scenarios and which contains the methodology and results of the subjective scoring for the data forming the basis of the present document.
  - French speech sentences as defined in Recommendation ITU-T P.501 [i.13] for wideband and English speech sentences as defined in Recommendation ITU-T P.501 [i.13] for narrowband.

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] ETSI ES 202 396-1: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[i.2] ETSI EG 202 396-2: "Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise; Part 2: Background Noise Transmission - Network Simulation - Subjective Test Database and Results".

[i.3] Recommendation ITU-T P.835: "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm".

[i.4] Recommendation ITU-T P.800: "Methods for subjective determination of transmission quality".

[i.5] Recommendation ITU-T P.831: "Subjective performance evaluation of network echo cancellers".

[i.6] Genuit, K.: "Objective Evaluation of Acoustic Quality Based on a Relative Approach", InterNoise '96, Liverpool, UK.

[i.7] Recommendation ITU-T SG 12 Contribution 34: "Evaluation of the quality of background noise transmission using the "Relative Approach"".

[i.8] ETSI 2nd Speech Quality Test Event: "Anonymized Test Report", ETSI Plugtests, HEAD acoustics, T-Systems Nova.

NOTE: Available at http://www.etsi.org/WebSite/OurServices/Plugtests/History.aspx. Also available as ETSI TR 102 648-3.

[i.9] ETSI 3rd Speech Quality Test Event: "Anonymized Test Report "IP Gateways".

NOTE: Available at http://www.etsi.org/WebSite/OurServices/Plugtests/History.aspx.

[i.10] ETSI 3rd Speech Quality Test Event: "Anonymized Test Report "IP Phones".

[i.11] ETSI 4th Speech Quality Test Event: "Anonymized Test Report "IP Gateways and IP Phones".

NOTE: Available at http://www.etsi.org/WebSite/OurServices/Plugtests/History.aspx.

[i.12] F. Kettler, H.W. Gierlich, F. Rosenberger: "Application of the Relative Approach to Optimize Packet Loss Concealment Implementations", DAGA, March 2003, Aachen, Germany.

[i.13] Recommendation ITU-T P.501: "Test Signals for Use in Telephonometry".

[i.14] R. Sottek, K. Genuit: "Models of Signal Processing in human hearing", International Journal of Electronics and Communications (AEÜ) volume 59, 2005, p. 157-165.

NOTE: Available at http://www.elsevier.de/aeue.

[i.15] SAE International - Document 2005-01-2513: "Tools and Methods for Product Sound Design of Vehicles" R. Sottek, W. Krebber, G. Stanley.

[i.16] Recommendation ITU-T P.862: "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs".

[i.17] Recommendation ITU-T P.862.1: "Mapping function for transforming P.862 raw result scores to MOS-LQO".

[i.18] Recommendation ITU-T P.862.2: "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs".

[i.19] Recommendation ITU-T SG 12 Contribution 19: "Results of objective speech quality assessment of wideband speech using the Advanced TOSQA2001".

[i.20] Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".

[i.21] Recommendation ITU-T G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[i.22] Recommendation ITU-T P.56: "Objective measurement of active speech level".

[i.23] Recommendation ITU-T P.57: "Artificial ears".

[i.24] M. Spiegel: "Theory and problems of statistics", McGraw Hill, 1998.

[i.25] Void.

[i.26] M. Kendall: "Rank correlation methods", Charles Griffin & Company Limited, 1948.

[i.27] Sottek, R.: "Modelle zur Signalverarbeitung im menschlichen Gehör", PHD thesis RWTH Aachen, 1993.

[i.28] Recommendation ITU-T P.830: "Subjective performance assessment of telephone-band and wideband digital codecs".

[i.29] Void.

[i.30] ANSI S1.1-1986 (ASA 65-1986): "Specifications for Octave-Band and Fractional-Octave-Band Analog and Digital Filters", 1993.

[i.31] Recommendation ITU-T G.160 Appendix II, Amendment 2: "Voice enhancement devices: Revised Appendix II - Objective measures for the characterization of the basic functioning of noise reduction algorithms".

[i.32] ETSI TS 103 106: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods".

[i.33] Hastie T.; Tibshirani R. and Friedman J.: "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", New York: Springer-Verlag, 2001.

[i.34] ETSI EG 202 396-3 (V1.1.1 to V1.3.1): "Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise; Part 3: Background noise transmission - Objective test methods".

3 Symbols and abbreviations

3.1 Symbols

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

σ^2 Variance

3.2 Abbreviations

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

AMR Adaptive MultiRate
ASL Active Speech Level

NOTE: According to Recommendation ITU-T P.56 [i.22].

BGN BackGround Noise
CDF Cumulative Density Function
dB SPL Sound Pressure Level re 20 µPa in dB
DB Data Base
DUT Device Under Test
EFR Enhance Full Rate
FR Full Rate
G-MOS Global MOS

NOTE: MOS related to the overall sample.

GSM Global System for Mobile Communication

HATS	Head And Torso Simulator
HiQ	High Quality (codec mode)
IP	Internet Protocol
IRS	Intermediate Reference System
ITU	International Telecommunication Union
ITU-T	Telecom Standardization Body of ITU
LQ	Low Quality (codec mode)
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MOS-LQSN	Mean Opinion Score - Listening Quality Subjective Noise
MRP	Mouth Reference Point
NB	NarrowBand
NBGN	Level of background noise
NI	Network I conditions
NII	Network II conditions
NIII	Network III conditions
N-MOS	Noise MOS

NOTE: MOS related to the noise transmission only.

NR	Noise Reduction
NR (filter)	Noise Reduction (filter)
NSA	Noise Suppression Algorithm
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
RCV	ReCeiVe
RMS	Root Mean Square
RMSE	Random Mean Square Error
SG	Study Group
S-MOS	Speech MOS

NOTE: MOS related to the speech signal only.

SND	Sending Direction
SNR	Signal to Noise Ratio
SQTE	Speech Quality Test Event
SPL	Sound Pressure Level
STD	STandard Deviation
STF	Specialized Task Force
TMOS	TOSQA Mean Opinion Score
TOR	Terms Of Reference
VAD	Voice Activity Detection
VoIP	Voice over IP
WB	WideBand

4 Speech signals to be used

As with any objective model, the prediction of speech quality depends on the conditions under which the model was tested and validated (see clauses 6.1 and 8). This dependency also applies to the speech material used in conjunction with the objective model.

The wideband version of the model uses French speech sentences. The near end speech signal (clean speech signal) consists of 8 sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be taken from Recommendation ITU-T P.501 [i.13].

The narrowband version of the model uses English speech sentences. The near end speech signal (clean speech signal) consists of 8 sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be taken from Recommendation ITU-T P.501 [i.13].

5 Selection of the data within the scope of the wideband objective model: Experts evaluation

5.1 Selection process

The aim of the selection process was to identify those data in the databases described in ETSI EG 202 396-2 [i.2] which are consistent with the scope of the objective models to be studied within the present document.

The experts were selected on the based on the definition found in e.g. Recommendation ITU-T P.831 [i.5]: experts are experienced in subjective testing. Experts are able to describe an auditory event in detail and are able to separate different events based on specific impairments. They are able to describe their subjective impressions in detail. They have a background in technical implementations of noise reduction systems and transmission impairments and do have detailed knowledge of the influence of particular implementations on subjective quality.

Their task was to select the relevant conditions within the scope of the model to be developed. Therefore they had to verify the consistency of the data with respect to the following selection criteria:

1) Artefacts others than the ones which should have been produced by the signal processing described in ETSI EG 202 396-2 [i.2] e.g. due to the additional amplification required in order to provide a listening level of 79 dB SPL.

2) Inconsistencies within one condition due to the selection of the individual speech samples from the database for subjective evaluation.

3) Inconsistencies within one condition due to statistical variation of the signal processing described in ETSI EG 202 396-2 [i.2] leading to non consistent judgements within this condition.

4) Inconsistencies due to Recommendation ITU-T P.56 [i.22] level adjustment process chosen for the complete files including the background noise.

As a result of the experts listening test a set of data was selected which is used for the development of the objective model.

In the selection process five expert listeners (non-native French speakers) were involved. Their task was not to produce new judgements, but to check all the samples in the database with respect to the possible artefacts described above.

A playback system with calibrated headphones was used for the test. The equalization provided by the headphone manufacturer was used since this was the one used in the auditory French test setup.

NOTE: These headphones and headphone amplifiers were used in the tests since they provide the performance required. Other products providing the equivalent performance could be used if such an experiment should be repeated by others. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of these products.

All samples could be heard by the experts as often as required in order to get final agreement about the applicability of the data within the terms of reference of the model. There was no limitation in comparing samples to the ones previously heard.

5.2 Results

In general it could be observed that the 4 seconds sample size chosen in the experiment according to Recommendation ITU-T P.835 [i.3] lead to a more difficult task even for expert listeners, especially in the case of non-stationary background noises. It is more difficult to identify the nature of the noise itself and then identify in addition possible impairments introduced by the signal processing or by the network impairments. It is very likely that some comparatively high standard deviations seen in the data are caused by these effects.

5.3 French database

In general the French database is in line with the ToR except network condition NII. In network condition NII 1 % packet loss was chosen which is too low for the conditions to be evaluated. Due to the inhomogeneously distributed packet losses there are conditions where no packet loss is audible up to conditions where 5 out of 6 samples show packet loss. Furthermore the packet loss may occur during speech as well as during the noise periods. The impact of the different packet losses is not controlled with respect to their occurrence due to the statistical nature of the packet loss distribution, even within a set of 6 samples used for evaluating one condition. Since packet loss is clearly audible under NIII conditions (3 % packet loss) and much better distributed amongst the different samples the NII conditions are not used within the scope of the objective method. They are either covered by the NI condition (0 % packet loss) or by the NIII conditions. This results in 144 NII conditions which are not retained for the development of the model.

From the 288 NI and NIII conditions 28 conditions are not retained. The main reasons therefore are:

- Not consistent signal levels due to the amplification process.
- Insufficient S/N, speech almost inaudible.

The individual reasons for the samples of these conditions being not retained can be found in table A.1.

In total 260 out of 432 conditions are used as the reference for the objective model. In other words, 60,2 % of the data can be used for the model. The distribution of the ratings is between 1,2 and 4,96 MOS for S-/N-/G-MOS.

6 Description of the wideband objective test method

6.1 Introduction

The present objective test method is developed in order to calculate objective MOS for speech, noise and the overall quality of a transmitted signal containing speech and background noise, designated N-MOS, S-MOS and G-MOS in the following.

The new model is based on an aurally-adequate analysis in order to best cover the listener's perception based on the previously carried out listening test ETSI EG 202 396-2 [i.2].

The wideband objective model is applicable for:

- wideband handset and wideband hands-free devices (in sending direction);
- noisy environments (stationary or non-stationary noise);
- different noise reduction algorithms;
- AMR Recommendation ITU-T G.722.2 [i.21] and Recommendation ITU-T G.722 [i.20] wideband coders;
- VoIP networks introducing packet loss.

NOTE 1: For the NIII conditions jitter was introduced. Finally jitter was observed for less than 2 % of the selected conditions. The jitter consideration of the new objective method could therefore not be validated on an appropriate amount of data. Quality impairments typically introduced by different strategies of packet loss concealment and different adaptive jitter buffer control mechanisms were not considered in the listening test database and therefore also not in the objective method.

NOTE 2: The method is not applicable for such background situations where speech intelligibility is the major issue.

Due to the special sample generation process the new method is only applicable for electrically recorded signals. The quality of terminals can therefore only be determined in sending direction.

The method was developed by attaching importance to a high reliability. The results of the listening test (selected conditions, see clause 5) were best modelled. Furthermore mechanisms were implemented to provide high robustness also for other than the present samples.

The sample preparation and nomenclatures for the new method are described in clause 6.2.

The calculation of *N-MOS*, *S-MOS* and *G-MOS* is described in detail in clauses 6.5 to 6.7.

6.2 Speech sample preparation and nomenclature

6.2.1 Speech sample preparation

Based on the data selected in clause 5 an objective model is developed in order to determine:

- the Noise-MOS (N-MOS);
- the Speech-MOS (S-MOS); and
- the "Global"-MOS (G-MOS), the overall quality including speech *and* background noise.

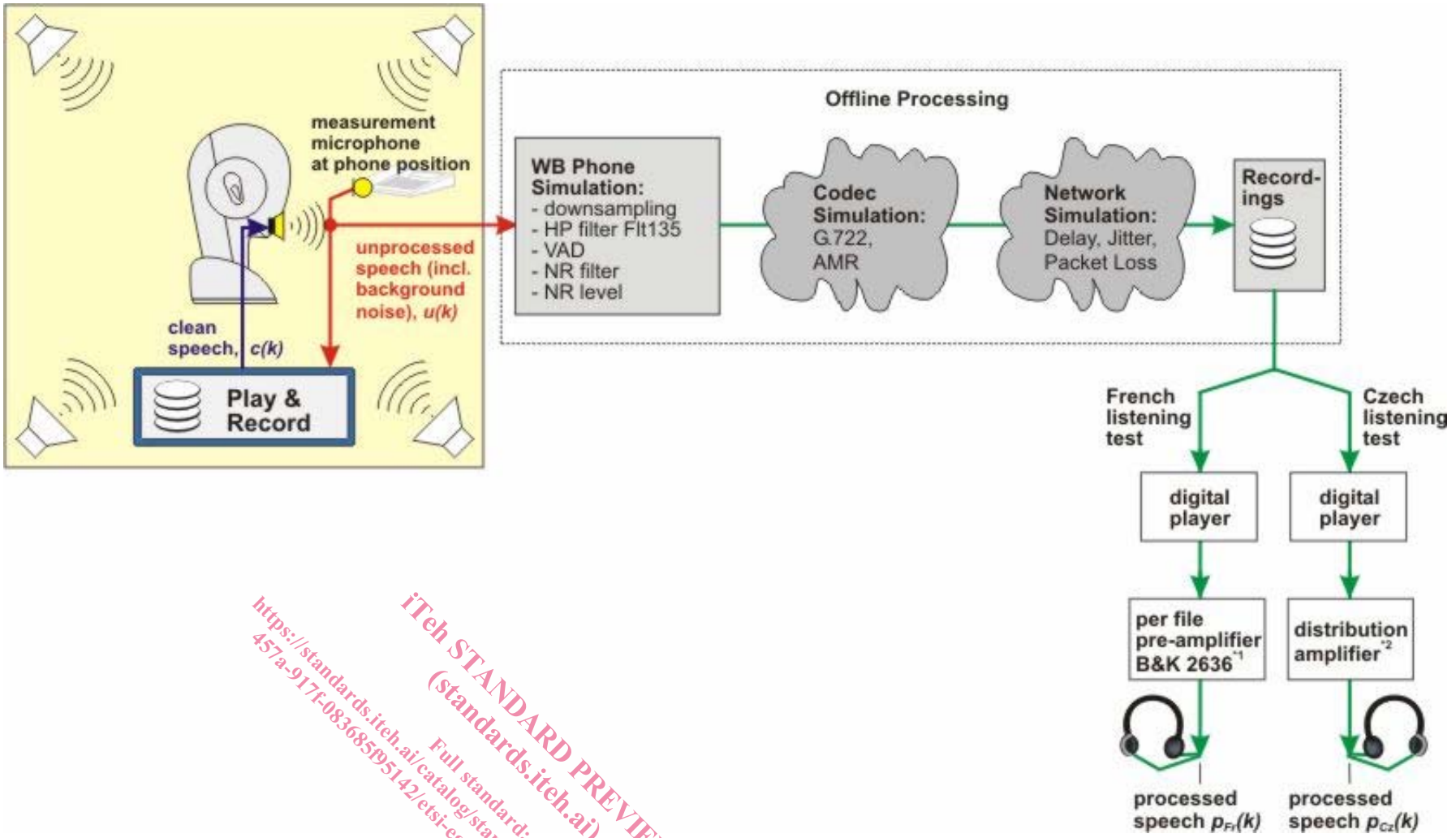
Different input signals can be accessed during the recording process and subsequently can be used for the calculation of N-MOS, S-MOS and G-MOS. Beside the signals used in the listening test ("processed signal"), two additional signals are used as a priori knowledge for the calculation:

1) The "clean speech" signal, which was played back via the artificial mouth at the beginning of the sample generation process.

2) The "unprocessed signal", which was recorded close to the microphone position of the simulated handset device/hands-free telephone (see figure 6.1 and ETSI EG 202 396-2 [i.2]). Note that no real phone/hands-free device was used. Phones and handsfree devices were simulated by a free-field microphone and an offline simulation for filtering, VAD, noise reduction, etc.

Both signals are used in order to determine the degradation of speech and background noise due to the signal processing as the listeners did during the listening tests.

The sample generation process is shown in figure 6.1.



NOTE 1: Calibrated for each file with B&K HATS (3.3 ears) to 79 dB SPL ASL (Recommendation ITU-T P.56 [i.22]).
NOTE 2: Once calibrated: -26 dBoV resulting to 79 dB SPL measured with a type 3.2 ear (Recommendation ITU-T P.57 [i.23]), 5N application force.

Figure 6.1: Sample generation process, indicating "clean speech", "unprocessed speech" and "processed speech"