ETSI EG 202 396-2 V1.1.1 (2006-09)

ETSI Guide

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

Reference

DEG/STQ-00038-2

Keywords

noise, QoS, 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

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the 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
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

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://webapp.etsi.org/IPR/home.asp).

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 ETSI Guide (EG) has been produced by ETSI Technical Committee Speech Processing, Transmission and Quality Aspects (STQ).

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

The present document is part 2 of a multi-part deliverable covering 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".

1 Scope

The present document aims at setting up and verifying a transmission network simulation environment using realistic network scenarios for laboratory use in the context of background noise transmission in WideBand (WB) audio conversational applications.

Background noise is a problem in mostly all situations and conditions and needs to be taken into account in both terminals and networks. The document provides information about the transmission network impairments in packet based communication network since these effects tend to aggravate the consequences of background noise and need to be considered carefully. The present document includes:

- Setup of simulation environment (network and signal processing in terminals) providing comparable network characteristics and traffic patterns as found in conditions of the real transmission networks, based on general information on network impairment types and codec features.
- Description of an example of network simulation database containing the results of applying some typical and realistic transmission network scenarios and traffic patterns in a selected variety of environments.
- Description of how to produce a speech sample database using the setup given, and of the appropriate method to collect the corresponding subjective scores, for which an example of results, obtained for the purpose of ETSI STF 294, is given.

The setup and the process mentioned above are meant to be applied on speech samples. Nevertheless, although the resulting speech database is described in the present document (see clause 8), its production is outside the scope of the present document.

The setup, network simulation database and subjective test results as described in the document are applicable for:

- Simulation of network impairments in general (concerns the setup only).
- Evaluation of the contribution of background noise performance of terminals and networks to the perceived overall quality.
- Development of objective method for the quantification of background noise transmission performance (as it is developed in EG 202 396-3).

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 and/or edition number or version number) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.

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

[1] ITU-T Recommendation G.1010: "End-user multimedia QoS categories".

[2] ETSI TR 101 329-1: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 1: General aspects of Quality of Service (QoS)".

[3] ETSI TR 101 329-6: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 6: Actual measurements of network and terminal characteristics and performance parameters in TIPHON networks and their influence on voice quality".

[4] ITU-T Recommendation G.114: "One-way transmission time".

[5] ETSI TR 101 329-7: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 7: Design guide for elements of a TIPHON connection from an end-to-end speech transmission performance point of view".

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

[7] ITU-T Recommendation G.726: "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

[8] ETSI TS 126 171: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; General description (3GPP TS 26.171 version 6.0.0 Release 6)".

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

[10] ETSI TS 126 071: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech Codec; General description (3GPP TS 26.071 Release 6)".

[11] ETSI EN 300 726: "Digital cellular telecommunications system (Phase 2+) (GSM); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 Release 1999)".

[12] ITU-T Recommendation G.729: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)".

[13] ITU-T Recommendation G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s".

[14] ETSI TS 126 191: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; Error concealment of lost frames (3GPP TS 26.191 Release 6)".

[15] ETSI TS 126 192: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Mandatory Speech Codec speech processing functions AMR Wideband Speech Codec; Comfort noise aspects (3GPP TS 26.192 Release 6)".

[16] ETSI TS 126 193: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Source controlled rate operation (3GPP TS 26.193 Release 6)".

[17] ETSI TS 126 194: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Mandatory Speech Codec speech processing functions AMR Wideband speech codec; Voice Activity Detector (VAD) (3GPP TS 26.194 Release 6)".

[18] ETSI TS 126 201: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; Frame structure (3GPP TS 26.201 Release 6)".

[19] ITU-T Recommendation G.722.1: "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[20] IETF RFC 1890: "RTP Profile for Audio and Video Conferences with Minimal Control".

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

[22] ETSI EG 202 396-1: "Speech processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise Part 1: Background noise simulation technique and background noise database".

[23] ITU-T Recommendation P.341: "Transmission characteristics for wideband (150-7000 Hz) digital hands-free telephony terminals".

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

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

[26] Mattila, V.: "Objective Measures for the Characterization of the Basic Functioning of Noise Suppression Algorithms", MESAQIN 2003, Measurement of Speech and Audio Quality in Networks, May 2003, Prague, ISBN: 80-01-02822-4, pp. 5-15.

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

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

[29] ITU-T Recommendation H.245: "Control protocol for multimedia communication, annex O".

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

[31] ITU-T Recommendation P.58: "Head and torso simulator for telephonometry".

3 Abbreviations

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

ABR	Average Bit Rate
ACELP	Algebraic Code Excited Linear Prediction coder
AMR-NB	Adaptive Multi-Rate-NarrowBand
AMR-WB	Adaptive Multi-Rate-WideBand
CBR	Constant Bit Rate
CELP	Code Excited Linear Prediction
CI	Confidence Interval
CNG	Comfort Noise Generator
DSL	Digital Subscriber Line
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission
EFR	Enhanced Full Rate
FSF	Free Software Foundation
GSM	Global System for Mobile communications
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
LPC	Linear Predictive Coding
MLT	Modulated Lapped Transform
MOS	Mean Opinion Score
NB	NarrowBand
NIST	National Institute of Standards and Technology
NSA	Noise Suppressing Algorithm
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RTP	Real Time Protocol
SB-ADPCM	Sub Band-Adaptive Differential Pulse Code Modulation
SCR	Source Controlled Rate
SID	SIlence Description
SPL	Sound Pressure Level
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
VAD	Voice Activity Detection
VBR	Variable Bit Rate
VoIP	Voice over IP
WAV	WAVeform
WB	WideBand

WCDMA	Wideband Code Division Multiple Access
xDSL	x Digital Subscriber Line

4 Performance parameters

4.1 Overview

The speech quality (as well as the background noise transmission quality) when it is received at listener's side depends strongly on a good transmission network. There are lots of conditions and parameters that can influence on speech quality, being of difference grades of importance according to the network technology over which the audio signal is transmitted.

This clause aims at identifying the most relevant network parameters that could be considered as main degraders of speech quality. They have been selected focusing on those networks where WideBand (WB) codecs are currently being used, all of them packet networks.

Other parameters impacting speech or noise transmission quality, but not related to the network it self, are not addressed here.

4.2 Performance key parameters

4.2.1 Delay

One of the main Quality of Service (QoS) factors in voice transmissions is the delay perceived by the users. In general, delay denotes the amount of time it takes for a signal to reach a destination. It has a very direct impact on user satisfaction depending on the application, and manifests itself in many different ways. It includes delays in the terminal, network, and any servers. Delay also takes into account the effect of other network parameters such as bandwidth as can be seen in clause 4.3.

A great deal of delay may cause unavailability and non-consistent real-time interaction, with frustrating consequences for the conversation. In order to allow a normal conversation over a network, this delay has to be kept almost constant and below a defined limit. If the end-to-end delay is too high, an interactive communication is difficult or impossible.

In the context of ITU (International Telecommunication Union) the preferred delay is < 150 ms and the limit is considered < 400 ms (see ITU-T Recommendation G.1010 [1]). Some other studies in the scientific literature have led to the following conclusions (see TR 101 329-1 [2]):

- Delays lower than 10 ms to 15 ms are not annoying for users, since they do not perceive the echo effect.
- Delays up to 150 ms require echo control but do not compromise the effective interaction between the users.
- If the delays are in the range 200 ms to 400 ms, the effectiveness of the interaction is lower but can be still acceptable.
- A delay higher than 400 ms makes the interactive voice communication quite difficult and conversation rules are required.

End-to-end delay is the sum of several factors as presented in figure 1. Some of them are due to terminal equipment (such as codec delay), and others are due to the network. The contribution of some of these factors is described in clauses 4.2.1.1 to 4.2.1.6.

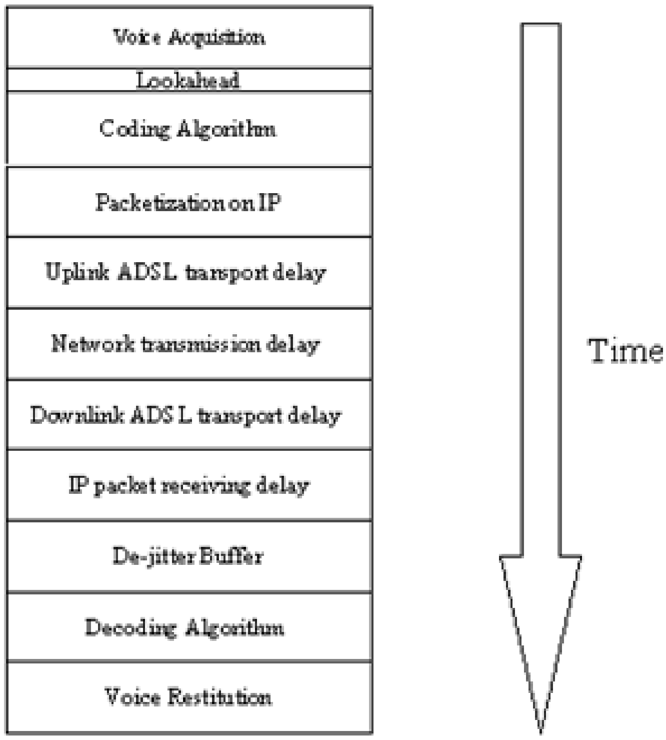


Figure 1: Example of a VoIP phone2phone communication with ADSL on both sides

4.2.1.1 Codec delay

It is the time taken by the compression on sending side (coding) as well as by the decompression on the receiving side (decoding). Because different codecs work in different ways, this delay varies with the voice codec used and process speed.

The compression algorithm must have some knowledge of what is in the next block to process in order to accurately reproduce the current block that is being processed. This look-ahead is an additional delay called algorithm delay and effectively increases the length of the compression block.

4.2.1.2 Packetization delay

The packetization delay is caused by the amount of time it takes to accumulate the bytes of a packet of a reasonable size. This delay is a function of the sample block size required by the codec and the number of blocks placed in a single frame. Generally, the larger the packet size, the greater the amount of time it takes to fill it.

Packetization delay can also be called accumulation delay since the voice samples are accumulated in a buffer before being release. As a general rule, this kind of delay should not exceed 30 ms.

4.2.1.3 Output queuing delay

After the compressed voice payload is built, a header is added and the frame is queued for transmission on the network connection. Output queuing delay is the amount of time a packet resides in the output queue of that interface. It is variable delay and is dependent on the trunk speed and the state of the queue.

4.2.1.4 Serialization delay

Another source of delay is the serialization of the digital data onto the physical media. It is the time it takes to put packets onto a link, and it is determined by packet size divided by the capacity of the link. This value is somewhat dependent on the link technology used and its access method. Although this delay is unavoidable, keeping the number of intervening links small and using high bandwidth interfaces reduces the overall latency, so it can be negligible in some cases.

4.2.1.5 Network delay

Network delay corresponds to the time it takes for packets to be carried by the network from the source to the destination. Network delay is caused by the combination of network propagation delay, processing delays, and variable queuing delays at the intermediate routers on the path to the destination terminal.

4.2.1.5.1 Network switching delay

This delay comprises a fixed delay and queuing delay.

Fixed switching delay or processing delay is the time it takes a network device to buffer packets and make the decision on which interface the packet is to be directed. Switching delays on today's high performance routers are negligible, typically in the order of 10 µs to 20 µs per packet. Although this delay is small, the architecture of the router or switch is the deciding factor.

Queuing delay, which is a large source of latency, is defined as the time difference between the enqueuing of a packet on the outbound interface scheduler, and the start of clocking the packet onto the outbound link. This is a function of the scheduling algorithm used and of the scheduler queue utilization, which is in turn a function of the queue capacity and the offered traffic load and profile.

4.2.1.5.2 Propagation delay

It is the amount of time it takes the signal to traverse the length of the specific medium of transmission. There is always propagation delay; however, it only becomes an important issue when the packets travel a great distance, specifically, this delay is usually negligible if links are shorter than 1 000 km (see TR 101 329-6 [3]). Therefore this latency is function of the distance and the kind of the physical link (e.g. 5 µs/km in optical fibre cable system - see ITU-T Recommendation G.114 [4]).

4.2.1.6 De-jitter delay

Jitter is delay variation. Because speech is a constant bit-rate service, the jitter must be removed before the signal leaves the network. The de-jitter buffer transforms the variable delay into a fixed delay, by holding the first sample received for a period of time before playing it out.

4.2.2 Jitter

Jitter characterizes the variation of delay. It is generally computed as the variation in the components of the delay previously described, for two consecutive packets. It is usually caused by the buffers built up on routers during periods of increased traffic. Propagation delay can vary as network topology changes, when a link fails, for example, or when the topology of a lower layer network changes, causing a sudden peak of jitter. Switching delay can also vary as some packets might require more processing than others and variation in queuing delay is caused as scheduling queues oscillate between empty and full. This variability in the delay also creates the possibility of asymmetric links, in which delays may be different in the two directions of the conversation.

Delay variation is generally included as a performance parameter due to the inherent variability in arrival times of individual packets. However, services that are highly intolerant of delay variation will usually take solutions to remove, or significantly reduce, the delay variation by means of buffering, although at the expense of adding additional fixed delay. The buffers used to remove delay variation are denominated de-jitter buffers, and their duty is turning variable network delays into constant delays at the destination end systems. De-jitter buffer size is usually set to the maximum expected jitter. Packets delayed by more than that are considered as lost and default content is filled in. This may lead to audio defects (noise). Still, there are limits to buffering; these limits depend on the delay tolerance of the application and on buffer memory constraints.

In the context of ITU, the performance target for audio applications is jitter < 1ms (see ITU-T Recommendation G.1010 [1]). This value is meant after the de-jitter buffer. Otherwise, a jitter of up to about 30 ms can be dealt by any de-jitter buffer and does not cause any problematic supplementary delay.

4.2.3 Packet loss

Packet loss is typically the result of congestion in the transport network. Packet loss rate is defined as the percentage of data packets, out of the total number of transmitted packets that are lost somewhere along the path from the source to the destination.

A common scenario for packet loss is when an output queue in a network element is full of packets to be transmitted where there is no additional memory space for additional ingress packets; this condition is commonly referred to as an output queue full condition. In this condition, the network device that is queuing the packets has no choice but to drop the packet. For example, an output queue full condition can occur where a sender attached to an interface of a higher speed is sending to a receiver attached to an interface of a lower speed. Eventually, the output queue buffers become full, resulting in dropped packets. However information loss is not limited to the effects of bit errors or packet loss during transmission, but also includes the effects of any degradation introduced by media coding for more efficient transmission.

Information loss has a very direct effect on the quality of the information finally presented to the user, whether it is voice, image, video, or data. In the context of ITU, the performance target for audio applications is a packet loss ratio < 3 % (see ITU-T Recommendation G.1010 [1]).

4.3 Parameter interaction and dependences

The influence of the main transmission parameters on perceived quality is shown in the figure 2:

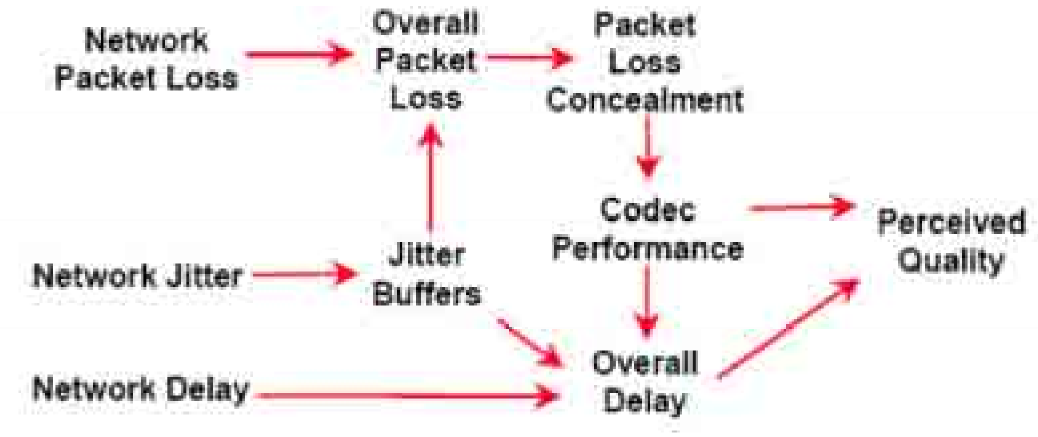


Figure 2: Interaction of transmission aspects (see TR 101 329-7 [5])

These parameters are not independent one another, but they have strong dependences. The relationships are not bidirectional. In case a parameter varies, this fact has implications into another one, but if the latter changes, the first may change or not.

- Because the maximum jitter is the maximum difference between measured delays, an increase in the delay will raise the probability of increasing jitter. Generally jitter variation does not modify the delay, however, the higher the jitter is, the larger the de-jitter buffer size has to be configured, contributing to increase the whole end-to-end delay.
- Long delays can cause some packets to be discarded, increasing packet loss probability. However, delays do not go up because of more packet losses.
- Jitter is also associated with the loss or de-sequencing of data packets in a real-time data stream. This is why an increase in the packet loss rate contributes to increase the jitter. However the jitter does not influence the packet loss rate in the network.

5 Description of codecs features

One will find below some features of codecs that can be quoted amongst the main ones.

5.1 Speech coding algorithm

5.1.1 Waveform codecs

Waveform coders (sinusoidal, etc.) are characterized by their attempt to preserve the general shape of the signal waveform.

5.1.2 Vocoders

At the opposite extreme to waveform coders, vocoders (LPC, etc.) are very speech specific in their principles as no attempts are made to preserve the original speech waveform, consisting of an analyzer and a synthesizer.

5.1.3 Hybrid codecs

To overcome the deficiencies of pure waveform and vocoding schemes, hybrid coding (PCM, CELP, etc.) methods have been developed which incorporate advantages offered by each of the pure schemes. Hybrid coders broadly fall into two sub categories: frequency domain and time domain analysis.

5.2 Bit-rate

The bit-rate of a coder (expressed in kbit/s) is determined by the multiplication of its sampling frequency (expressed in Hz, generally 8 kHz) by its sample size (expressed in bits).

5.2.1 Constant Bit Rate (CBR)

CBR encoding means that the rate at which a codec's output data should be consumed is constant. CBR is useful for streaming multimedia content on limited capacity channels since it is the maximum bit rate that matters, not the average, so CBR would be used to take advantage of all of the capacity. CBR would not be the optimal choice for storage as it would not allocate enough data for complex clauses (resulting in degraded quality) while wasting data on simple clauses.

5.2.2 Variable Bit Rate (VBR)

VBR allows a codec to change quality over the course of the speech by using a quality rating based on changing original file quality; more data is used for increasing detail (e.g. sounds like vowels and high-energy transients), less for decreasing (e.g. fricatives sounds). For this reason, VBR can achieve better use of sound quality and file size.

Despite its advantages, VBR has some drawbacks:

- There is no guaranty about the final average bit-rate.
- Encoding using VBR can take considerably longer than using an equivalent fixed bit-rate.
- For some real-time applications like Voice over IP (VoIP), what counts is the maximum bit-rate, which must be low enough for the communication channel.
- It must be re-encoded in order to be streamed, involving a loss of quality.

Average Bit Rate (ABR) solves one of the problems of VBR, as it dynamically adjusts VBR quality in order to meet a specific target bit-rate, adding bits where necessary.

5.3 Discontinuous Transmission (DTX)

This is an addition to Voice Activity Detection (VAD)/VBR operation, and it is a method that suspends transmission in case a pause in the normal flow of conversation when is detected in the telephone device and background noise is stationary. DTX, also known as Source Controlled Rate (SCR), is composed of the VAD and Comfort Noise Generator (CNG) algorithms.

5.3.1 Voice Activity Detection (VAD)

It is used to reduce the transmission rate during silence periods of speech. The purpose of VAD is to detect whether the audio being encoded is speech. Two situations are possible:

- presence of CNG algorithm: VAD conveys the proper information to the CNG algorithm; or
- absence of CNG algorithm: non-speech periods are encoded with enough bits to reproduce the background noise.

VAD is always implicitly activated when encoding in VBR.

5.3.2 Comfort Noise Generator (CNG)

CNG algorithm allows the insertion of an artificial noise during silent intervals of speech. The purpose of the CNG algorithm is to create a noise that matches the actual background noise with a global transmission cost as low as possible.

5.4 Packet Loss Concealment (PLC)

Packet loss can be bursty in nature. The average packet loss rate for a call may be low; however, these periods of high loss rate can cause noticeable degradation in speech quality. PLC is a technique used to mask the effects of lost or discarded packets. PLC algorithms involve either replaying the last packet received or some more sophisticated algorithm that uses previous speech samples to generate speech. PLC is effective only for small numbers of consecutive lost packets and for low packet loss rates.

5.5 Perceptual enhancement

Perceptual enhancement is a set of optional post processing which can attempt to enhance the quality of the signal and to reduce the perception of the artefacts produced by the coding/decoding process. An example of such processing is bandwidth extension.

6 Wideband codecs

6.1 Overview

Wideband codecs increase intelligibility and naturalness of speech, create a feeling of transparent communication and facilitate speaker recognition, by using a wider band of frequency (50 Hz to 7 000 Hz) compared to traditional NarrowBand telephone speech (200 Hz to 3 400 Hz). The low-frequency enhancement from 50 Hz to 200 Hz contributes to increased naturalness, presence, and comfort. The high-frequency extension from 3 400 Hz to 7 000 Hz provides better fricative differentiation and higher intelligibility.

In order to cover the whole frequency range, WideBand (WB) speech signals must be digitalised with a sampling pace at 16 kHz. If each sample is represented by 16-bit integers, this results a raw bit-rate of 256 kbit/s. Thus, speech compression becomes of significant importance for most WideBand (WB) speech communication applications.

In this clause, a short description of several relevant WideBand (WB) speech codecs within standardization groups and industry (terminal manufacturers and speech application developers) and which have been used to complete the network transmission simulation is introduced.

6.2 Description of WideBand codecs

The two codecs used for the purpose of ETSI STF 294 (which drafted the present document) are ITU-T Recommendation G.722 [6] and AMR-WB. This does not imply any advantage (in terms of quality or any other consideration) for these codecs in comparison to others. This is why, in addition to the description of these two codecs, and for the benefit of the users of this ETSI Guide, other codecs are also presented.

6.2.1 ITU-T Recommendation G.722

6.2.1.1 Overview

ITU-T Recommendation G.722 [6], approved in 1988, is the benchmark coder for WideBand (WB) speech coding quality which may be used for a variety of higher quality speech applications.

The coding system uses Sub Band-Adaptive Differential Pulse Code Modulation (SB-ADPCM). This technique splits the frequency band into two sub-bands and the signals are encoded using ADPCM, utilizing techniques similar to those in the ITU-T Recommendation G.726 [7] NarrowBand standard.

This codec was designed for Integrated Services Digital Network (ISDN) and videoconferencing.

6.2.1.2 Modes of operation

The coding bit rate is 64 kbit/s and the system involving G.722 coder can be used to work in three modes: 64 kbit/s (mode 1), 56 kbit/s (mode 2), and 48 kbit/s (mode 3). The latter two modes will allow an auxiliary data channel of 8 kbit/s and 16 kbit/s respectively, within the 64 kbit/s channel. The ITU-T Recommendation G.722 [6] full duplex algorithm converts 256 kbit/s linear PCM to and from 64 kbit/s, 56 kbit/s, or 48 kbit/s. The system is referred as 64 kbit/s (7 kHz) audio coding.

6.2.1.3 Encoder and decoder

The encoder operates on speech frames of 10 ms, corresponding to 160 samples, each consisting of 16 bits, at a sampling rate of 16 kHz. Therefore, frame size is 320 bytes before encoding. Frame size after encoding is 80 bytes.

The ITU-T Recommendation G.722 [6] audio encoder includes, as represented in figure 3:

- a **transmit audio part:** which converts an audio signal to a uniform digital signal which is coded using 14 bits with 16 kHz sampling;
- a **SB-ADPCM encoder:** which reduces the bit rate to 64 kbit/s; and
- a **data insertion device:** which makes use of, when needed, 1 or 2 audio bits per octet depending on the mode of operation and substitutes data bits to provide an auxiliary data channel of 8 kbit/s or 16 kbit/s respectively.

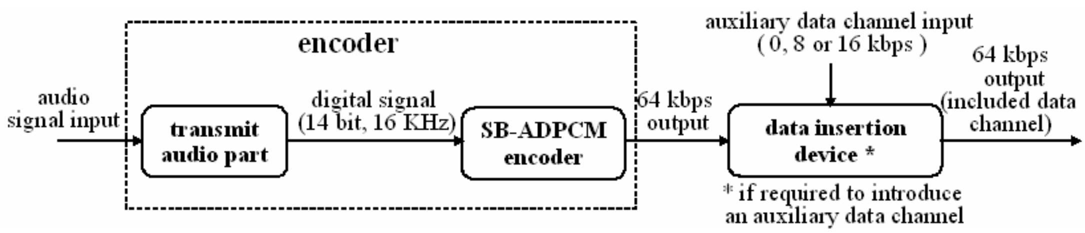


Figure 3: ITU-T Recommendation G.722 [6] audio encoder

The ITU-T Recommendation G.722 [6] audio decoder includes, as represented in figure 4:

- a **data extraction device:** which determines the mode of operation according to a mode control strategy and extracts the data bits as appropriate;
- a **SB-ADPCM decoder:** which performs the reverse operation to the encoder, noting that the effective audio coding bit-rate at the input of the decoder can be 64 kbit/s, 56 kbit/s, or 48 kbit/s depending on the mode of operation; and
- a **receive audio part:** which reconstructs the audio signal from the uniform digital signal which is encoded using 14 bits with 16 kHz sampling.

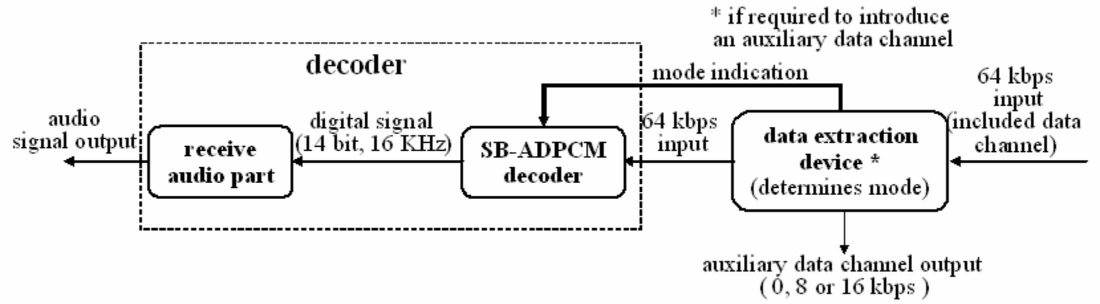


Figure 4: ITU-T Recommendation G.722 [6] audio decoder

6.2.1.4 Frame structure

ITU-T Recommendation G.722 [6] encoded frames are composed by speech and data channels according to:

- **mode 0:** 80 bytes of speech;
- **mode 1:** 70 bytes of speech and 10 bytes of data; and
- **mode 2:** 60 bytes of speech and 20 bytes of data.

6.2.2 AMR-WB

6.2.2.1 Overview

Adaptive Multi Rate (AMR) WB standard was approved by Third Generation Partnership Project (3GPP) / ETSI in 2001 (see TS 126 171 [8]). Subsequently, in 2003, it was named and approved Recommendation G.722.2 by ITU-T [9]. It was designed for Wideband Code Division Multiple Access (WCDMA) 3G and Global System for Mobile communications (GSM) systems, since advances in speech coding made WideBand (WB) coding feasible at the bit-rates applicable to mobile communication.

The AMR-WB speech codec utilizes the ACELP (Algebraic Code Excited Linear Prediction Coder) technology, which is employed also in the AMR-NB (see TS 126 071 [10]) and EFR (Enhanced Full Rate, see EN 300 726 [11]) speech codecs, as well as in ITU-T Recommendations G.729 [12] and G.723.1 [13] at 5.3 kbit/s.

Some of AMR-WB's capabilities are:

- error concealment mechanisms (see TS 126 191 [14]);
- CNG (see TS 126 192 [15]);
- SCR in 3G or DTX in GSM (see TS 126 193 [16]); and
- VAD (TS 126 194 [17]).

The foreseen applications for AMR-WB are the following: VoIP and Internet applications, mobile communications, PSTN (Public Switched Telephone Network) applications, ISDN WideBand (WB) telephony, ISDN video-telephony and video-conferencing.

Annexes A and B, and Appendix I of ITU-T Recommendation G.722.2 [9] provide supplemental functionalities allowing interoperability with GSM and 3GPP wireless systems. These functionalities have originally been developed for these systems, but their use is not limited to mobile applications.

6.2.2.2 Modes of operation

The AMR-WB codec is an adaptive codec capable of operating at nine speech coding bit rates: 6,6 kbit/s, 8,85 kbit/s, 12,65 kbit/s, 14,25 kbit/s, 15,85 kbit/s, 18,25 kbit/s, 19,85 kbit/s, 23,05 kbit/s, and 23,85 kbit/s.

The 12,64 kbit/s mode and the modes above offer high quality speech. The two lowest modes at 8,85 kbit/s and 6,6 kbit/s are intended to be used only temporarily during severe radio channel conditions or during network congestion.

The codec also includes a background noise mode designed to be used in the discontinuous transmission operation of GSM and, in other systems, as a low bit rate source-dependent mode for coding background noise. In GSM the bit rate of this mode is 1,75 kbit/s.

6.2.2.3 Encoder and decoder

Input to the encoder is a 16-bit PCM speech data sampled at 16 kHz. Each input speech frame of 20 ms consists of 320 16-bit PCM words containing 14-bit left-aligned uniform samples. The encoder outputs compressed speech data in octet aligned (by using bit stuffing) according to that defined in the TS 126 201 [18]. The decoder outputs the reconstructed speech data in the same format than input.

The AMR-WB audio encoder includes, as represented in figure 5:

- a **transmit audio part:** which converts an audio signal to a uniform digital signal which is coded using 14 bits with 16 kHz sampling;
- a **speech encoder:** which reduces the bit-rate depending on the codec mode;
- a **VAD:** which generates a flag indicating whether speech is being transmitted or not. This flag is an input of DTX control and operation module;
- a **comfort noise tx functions module:** which produces a frame of silence, shorter than that transmitted when voice is detected; and
- a **DTX control and operation module:** which has as inputs VAD flag and SIlence Description (SID) frame, and as outputs a 3-bit word that describes the type of information being transmitted and encoded speech frame itself.

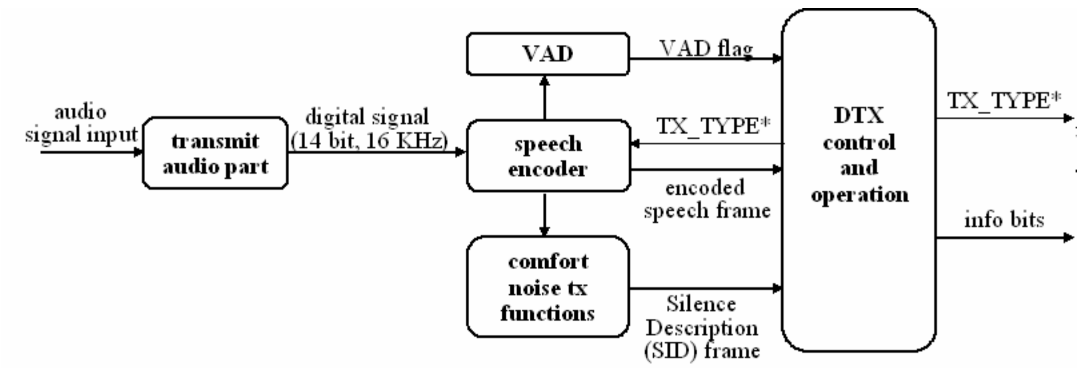


Figure 5: AMR-WB audio encoder

The AMR-WB audio decoder includes, as represented in figure 6:

- a **DTX control and operation module:** which extracts the encoded speech or SID frame from inputs received (information bits, type of frame and flags);
- a **speech decoder:** which performs the reverse operation to the encoder;

- a **speech frame substitution module:** with error concealment mechanisms in case of lost frames;
- a **comfort noise rx functions module:** which takes short silence frames and makes the inverse operation than in transmission so that speech decoder will be able to decode the signal correctly; and
- a **receive audio part:** which reconstructs the digital audio signal.

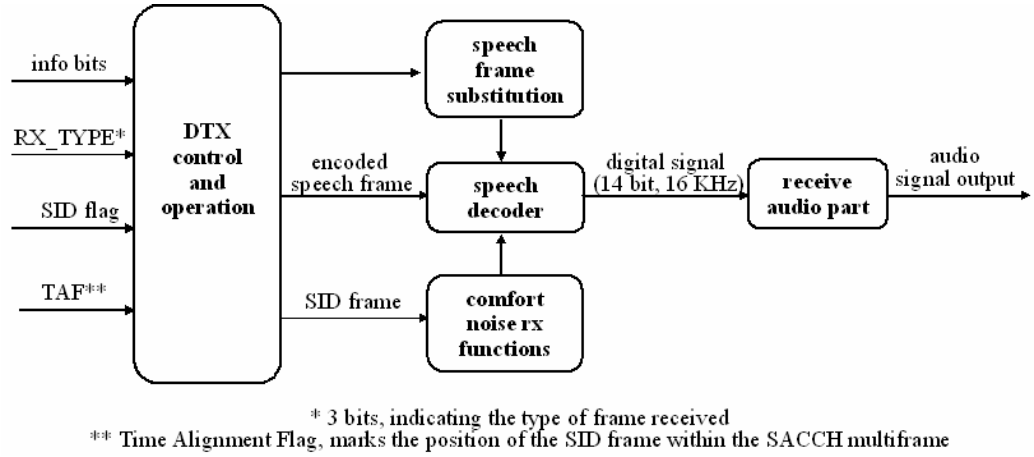


Figure 6: AMR-WB audio decoder

6.2.2.4 Frame structure

3GPP defines two frame formats for AMR-WB speech codec:

- AMR-WB interface format 1 (AMR-WB IF1); and
- AMR-WB interface format 2 (AMR-WB IF2), which includes octet alignment.

In this network simulation an implementation based on AMR-WB IF2 has been used, so, this is the one to be described in this clause.

The AMR-WB IF2 frame is formed by (see figure 7) the following fields:

- **frame type (4 bits):** identifying the current frame as either an AMR-WB codec mode, comfort noise, lost speech, or empty frame;
- **frame quality indicator (1 bit):** which indicates a bad of corrupted frame when equal to zero;
- **AMR-WB core frame:** which is the compressed speech data or comfort noise data; and
- **bit stuffing:** that completes the total number of bits in the frame to achieve an octet structure because it is typically not a multiple of eight.

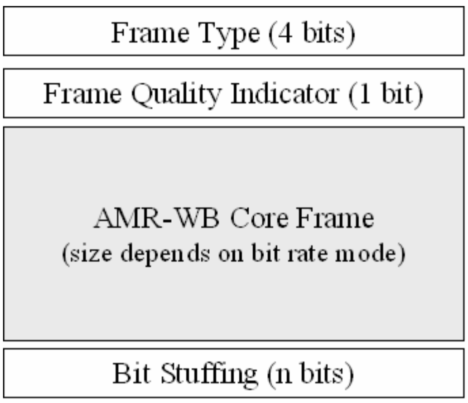


Figure 7: AMR-WB frame structure

Table 1 shows the composition of AMR-WB IF2 frames for all frame types in terms of how many bits are used for each field.

Table 1: AMR-WB frame composition

Frame Type Index	Bit rate (kbit/s)	Frame type (bits)	Frame quality indicator bits	AMR-WB core bits	Padding bits	Total bytes per AMR-WB IF2 frame
0	6,60	4	1	132	7	18
1	8,85	4	1	177	2	23
2	12,65	4	1	253	6	33
3	14,25	4	1	285	6	37
4	15,85	4	1	317	6	41
5	18,25	4	1	365	6	47
6	19,85	4	1	397	6	51
7	23,05	4	1	461	6	59
8	23,85	4	1	477	6	61
9	"SID" frame (comfort noise) (see note)	4	1	40	3	6
10-13	(for future use)	-	-	-	-	-
14	speech lost	4	1	0	3	1
15	no data	4	1	0	3	1
NOTE: Bit rate of comfort noise is 1,75 kbit/s when assuming continuous transmission.						

6.2.3 Other wideband codecs

6.2.3.1 Speex WB

6.2.3.1.1 Overview

Speex (see bibliography) is an open source patent-free audio compression format designed for speech. Speex is part of the GNU project and is not standardized but is currently quite used by industry.

The GNU Project (see bibliography) was launched in 1984 to develop a complete UNIX like operating system which is free software. The Free Software Foundation (FSF) (see bibliography) is the principal organizational sponsor of the GNU Project. FSF receives very little funding from corporations or grant-making foundations, but rely on support from individuals.

Speex is based on sub-band Code Excited Linear Prediction (CELP) modulation, because it has long proved that it could do the job and scale well to both low bit-rates and high bit-rates.

Some of Speex's features can be quoted:

- intensity stereo encoding;
- PLC;
- dynamic bit-rate switching and VBR operation;
- VAD, integrated with VBR;
- DTX; and
- variable complexity controlling how the search is performed with an integer ranking from 1 to 10. In practice, the best trade-off is between complexity 2 and 4, though higher settings are often useful when encoding non-speech sounds like Dual Tone Multi-Frequency (DTMF) tones.

Speex WB is well-suited to handle VoIP, Internet audio streaming, data archival (like voice mail), etc.

Designing for VoIP instead of cell phone use means that Speex must be robust to lost packets, but not to corrupted ones since packets either arrive unaltered or do not arrive at all. Also, the idea was to have a reasonable complexity and memory requirement without compromising too much on the efficiency of the codec.

6.2.3.1.2 Modes of operation

Speex WB codec is an adaptive codec capable of operating at multiple bit-rates from 2,15 kbit/s to 44 kbit/s. The bit rate is composed of narrow and WideBand (WB) bit-rates. The available NarrowBand (NB) bit-rates are 0,250 (no transmission, DTX), 2,15 kbit/s, 3,95 kbit/s, 5,95 kbit/s, 8 kbit/s, 11 kbit/s, 15 kbit/s, 18,2 kbit/s and 24,6 kbit/s, and for the additional band are 0, 250 kbit/s (no transmission, DTX), 1,8 kbit/s, 5,6 kbit/s, 9,6 kbit/s and 17,6 kbit/s.

6.2.3.1.3 Encoder and decoder

For WideWand (WB), the Speex approach uses a filter to split the band in two. The low band (0- to 4 kHz) is encoded with the NarrowBand (NB) mode and the high band (4 kHz to 8 kHz) is encoded by a different way. The input speech frame size for each encoder (narrow and WideBand (WB)) is 20 ms, corresponding to 160 16-bit PCM words, and it outputs compressed speech data in frames with size that depends on the mode of operation being used, that is independent for each of them. The resulting WideBand (WB) encoded frame is the addition of the two independent encoders.

All the encoders and decoders involved in this process include special modules for PLC, DTX and VAD capabilities. The evaluation results from such modules have an influence on the mode of operation to be used to encode the speech frame, producing lower or higher bit rates according to the instantaneous conditions.

6.2.3.1.4 Frame structure

For the WideBand (WB) mode, the entire NarrowBand frame is packed before the high-band is encoded, as described in figure 8. This means that a WideBand (WB) frame may be correctly decoded by a NarrowBand decoder with the only caveat that if more than one frame is in the same packet, the decoder will need to skip the high-band parts in order to sync with the bit-stream.

The Speex WB frame is formed by the following fields:

- **WideBand (WB) bit (0):** indicating that it is the first part (NarrowBand) of the frame;
- **NarrowBand frame type (4 bits):** identifying the mode of operation of NarrowBand part;
- **Speex WB (NB part) core frame:** which is the compressed 0 kHz to 4 kHz speech information**;**
- **WideBand (WB) bit (1):** indicating that it is the second part (WideBand (WB)) of the frame;
- **WideBand (WB) frame type (3 bits):** identifying the mode of operation of WideBand (WB) part; and
- **Speex WB (WB part) core frame:** which is the compressed 4 kHz to 8 kHz speech information.

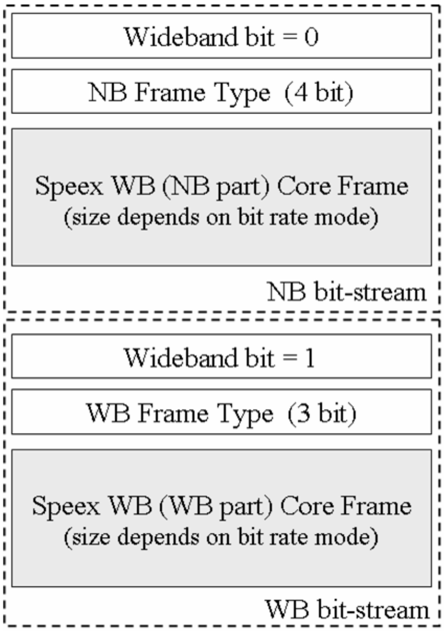


Figure 8: Speex WB frame structure

6.2.3.2 ITU-T Recommendation G.722.1

6.2.3.2.1 Overview

ITU-T Recommendation G.722.1 [19],standardized in 1999, operates at bit rates of 24 and 32 kbit/s with a better quality than ITU-T Recommendation G.722 [6] at 48 and 56, respectively, in a wide variety of operating conditions (clean speech, speech with background noise, music, tandeming, frame erasures, and level variation).

The algorithm is based on transform technology, using a Modulated Lapped Transform (MLT). It operates on 20 ms frames (320 samples) with a 20 ms of look-ahead buffer, resulting in a window size of 40 ms (640 samples). The transform coefficients are grouped in 500 Hz bands, and the gains in the bands are quantized and Huffman coded. The scaled coefficients are quantized using a perceptual categorization procedure and then Huffman coded.

There are two limitations of ITU-T Recommendation G.722.1 [19]. One is the lack of characterization tests and another one is the lack of work for developing a VAD/DTX/CNG algorithm for the standard.

6.2.3.2.2 Encoder and decoder

The digital input to the coder may be 14, 15 or 16 bit 2's complement format at a sample rate of 16 kHz (handled in the same way as in ITU-T Recommendation G.722 [6]). The analogue and digital interface circuitry at the encoder input and decoder output conforms to the same specifications described in ITU-T Recommendation G.722 [6].

6.2.3.2.3 Modes of operation and frame size

ITU-T Recommendation G.722.1 [19] operates at 24 and 32 kbit/s and the frame size for each is 60 bytes/frame and 80 bytes/frame, respectively.

6.2.3.3 ITU-T Recommendation G.729 annex J (also known as G.729 EV)

The main terms of reference of this coder are:

- multi-rate coder (8 kbits/s to 32 kbits/s, by steps of 2 kbits/s);
- baseline coder at 8 kbits/s: ITU-T Recommendation G.729 main [12];
- scalability between the different bit rates;
- operates in narrow band between 8 kbits/s and 12 kbits/s, and in wide band between 14 kbits/s and 32 kbits/s.

6.2.3.4 L16 (ITU-T Recommendation H.245 Annex O)

The L16 codec is an uncompressed audio data codec, using 16-bit signed representation with 65 535 equally divided steps between minimum and maximum signal level, ranging from -32 768 to 32 767. The value is represented in two's complement notation and network byte order. This codec is used for Acoustic Performance verification and possibly for low cost WideBand LAN applications. It is defined in the RFC 1890 [20] in clause 4.4.8.

6.3 Wideband codec comparison

Table 2 provides a comparison of WideBand (WB) codecs described above and their main features, in order to highlight the similarities and differences among all of them, be aware of the advantages and disadvantages of using one respect to the others and know the application which have been mainly designed for.

Table 2: Wideband codec comparison

Wideband codec	ITU-T Recommendation G.722 [6]	AMR-WB (ITU-T Recommendation G.722.2 [9])	Speex WB	ITU-T Recommendation G.722.1 [19]
Bit rate (kbit/s)	48, 56, 64	6,6 to 23,85	4 to 44,2	24, 32
Bandwidth	7 kHz	7 kHz	7 kHz	7 kHz/14 kHz
Multirate	No	Yes	Yes	No
Modulation	Sub-Band ADPCM	ACELP	CELP	Modulated Lapped Transform (MLT)
VAD	No	Yes	Yes	No
DTX	No	Yes	Yes	No
CNG	No	Yes	Yes	No
PLC	No	Yes	Yes	No
Main applications	ISDN, videoconference	Voice over IP (VoIP), Internet applications, Mobile Communications, 3G wireless, PSTN applications, ISDN WideBand (WB) telephony, ISDN video-telephony and video-conferencing	VoIP, Internet audio streaming, voice mail	ISDN, videoconference, VoIP
Status	Standardized by ITU-T	Standardized by ITU-T, ETSI/3GPP	Open-source	Standardized by ITU-T

7 Background noise transmission simulation

7.1 General description

This clause describes the simulation setups for WideBand (WB) terminals, network conditions, and their various combinations, that have been used to create the speech sample database containing typical and realistic transmission network scenarios. The final speech sample database contains speech samples affected by the most relevant impairments as well as speech enhancement processes from speaker to the output of the network. The work to be done for the simulation has been split in different steps as follows and as shown in figure 9:

1) Record clean speech sequences with a high quality system.

2) Add noisy conditions to the recordings.

3) Process the recorded noisy signals to simulate the behaviour of WideBand (WB) terminals.

4) Process the files with the network simulator.

5) Record speech sequences with the network impairments.

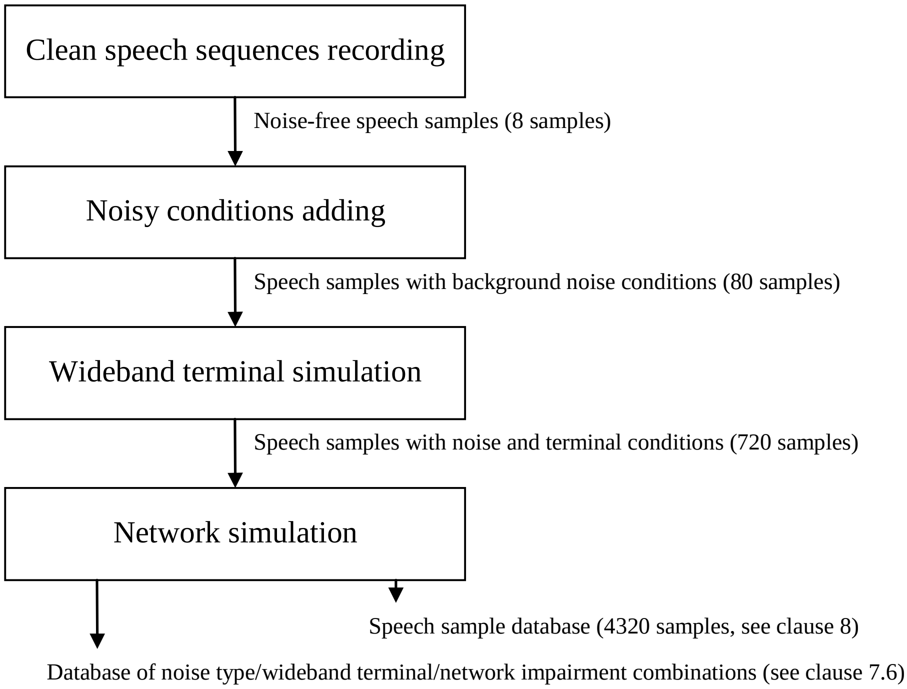


Figure 9: Background noise transmission simulation steps

These steps are analyzed in details in the following clauses. As for the choice of WideBand (WB) codecs in clause 6 above, the actual setup used for ETSI STF 294 will be provided as practical examples.

7.2 Speech sequences

The first phase of the process consists of recording a representative number of speech sequences without any type of background noise.

For the recordings performed by ETSI STF 294 the following conditions have been used:

- 48 kHz (16 bit) sampling rate;
- wave format; and
- the active speech level has been equalized to -26 dBov (see ITU-T Recommendation P.56 [21], no filtering applied).

Also for STF 294, the number of samples to have a representative database large enough for subjective and objective evaluation has been decided:

- 4 speakers (2 male, 2 female), 8 sentences each;
- two languages: Czech and French;
- the length of recordings varies between 24 s and 73 s. They contain neutral sentences of 2 s to 3 s length separated by pauses. Speech activity factor varies between 30 % and 60 %. This arrangement allows for easy later editing to any required format.

All this information is given as an example. But it is recommended to use the same sampling rate, active speech level and minimum numbers of speakers and languages.

7.3 Noisy conditions

This step is based on the input coming from ETSI STQ STF 273 where a database containing background noise samples has been made, as it is described in EG 202 396-1 [22].

In order to create a representative number of real life situations, in terms of noisy conditions, a certain number of different background noises need to been recorded for each speech file. For the purpose of SFT 294, this number has been set to 5 (what seems to be a reasonable minimum), and the following types have been selected:

- cafeteria noise;
- office room noise;
- road noise;
- crossroads; and
- car noise (car hands-free at 130 km/h).

Each combination of speech file plus background noise is recorded at two microphone-loudspeaker positions:

- typical handset microphone position (with loudness ratings adjusted to 7 dB); and
- hands-free microphone position (with loudness ratings adjusted to 11 dB).

All this resulted in a total of 80 speech samples for ETSI STF 294.

7.4 Noisy signal processing

After the background noise is added to the clean speech, the noisy signal must be processed to take into account the influence of the terminal. The processing includes:

- Convolution with impulse response of WideBand (WB) terminals, since we have a noisy-speech database independent of the terminal. It can be associated afterwards with any terminal we could get and measure in an anechoic room.
- Application of WideBand (WB) Noise-Suppression Algorithm (NSA), based on short-term spectral attenuation using Wiener filters. The algorithm can be tuned according to several parameters of interest:
  - level of noise-reduction (sort of threshold from which noise is reduced);
  - strength or "aggressivity" of the filter (smooth or sharp);
  - quality of residual noise; and
  - quality of the reconstruction of the speech signal.

In the following, the actual way this signal processing has been implemented for ETSI STF 294 is given.

The signal including speech plus background noise has been down-sampled (from 48 kHz to 16 kHz) and filtered out using band-pass filters (P341and Flt135, see figure 10). P341 is a band-pass filter which corresponds to the specification of the ITU-T Recommendation P.341 [23]. Actually, the filter available in the ITU-T Recommendation G.191 [24] has been used. Flt135 is a modified version of P341filter where the lower frequency of the band-pass is shifted from 50 Hz to 135 Hz. Such a pre-processing, which suppresses the frequency below 135 Hz, is justified since the signal-to-noise ratio is generally very low at these frequencies. In that sense, high-pass filtering consists in a basic noise reduction process widely used (alone or associated to time varying noise reduction).

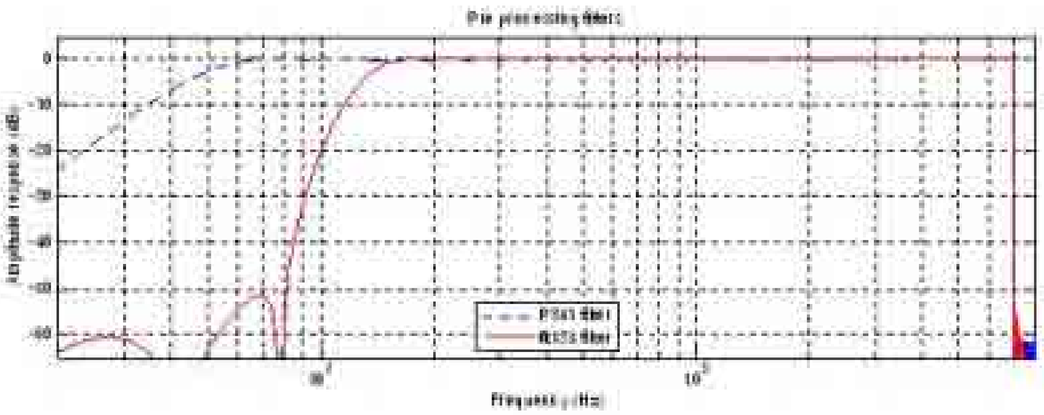


Figure 10: P341 and Flt135 band-pass filters

Then, the noise reduction algorithms are used to process these signals. The different algorithms are based on the same approach but with different tuning parameters and noise estimation procedure (3 parameters):

- Parameter 1:
  - with noise estimation using VAD; or
  - without VAD, e.g. continuous noise estimation.
- Parameter 2:
  - smooth noise reduction filter (small frequency resolution); or
  - sharp noise reduction filter (high frequency resolution).
- Parameter 3:
  - noise reduction level of 9 dB; or
  - noise reduction level of 18 dB.

Notice that pre-processing filters (P341and Flt135) are always used as noise reduction front-end.

Notice that real time implementation of each processing involves a delay that depends on the considered algorithm. In order to facilitate future works of objective evaluation, all processed files have been synchronized and have the same length.

All this resulted in a total of 1280 speech samples for ETSI STF 294.

7.5 Network simulation

7.5.1 General description

In this phase, noisy speech samples are simulated being transmitted over a network. This process adds delay, jitter, and packet loss. For the whole network simulation process a big black box is considered, where both input and output are wav signals and inside the functional blocks have to be defined and configured properly. These functional entities are:

- WideBand (WB) codecs to convert the wav signal into the proper format;
- RTP/UDP/IP transport modules;
- two IP soft-phones working as call generators, being in charge of transmitting and receiving the IP packets containing the WideBand (WB) samples; and
- an IP network emulator installed in a Linux PC with two IP addresses.

In order to generate the IP network conditions, a real-time network emulator must be used. NIST Net (see bibliography) has been chosen by the ETSI STF 294 and is recommended. This is a network emulator based on Linux and it is developed by the National Institute of Standards and Technology (NIST). NIST Net allows a normal PC to act as a router emulating a wide variety of network conditions. It is implemented as a module extension to the Linux kernel and is distributed as open source software. The NIST Net network emulator is a general-purpose tool for emulating performance dynamics in IP networks. The tool is designed to allow controlled, reproducible experiments with network performance sensitive/adaptive applications and control protocols in a simple laboratory setting. By operating at the IP level, NIST Net can emulate the critical end-to-end performance characteristics imposed by various wide area network situations (e.g. congestion loss) or by various underlying sub-network technologies (e.g. asymmetric bandwidth situations of xDSL and cable modems).

For the test laboratory the scenario in figure 11 should be setup.

Figure 11: Network simulation test-bed

The procedure for simulation is as follows:

1) The call generators establish a call.
2) WAV files are encoded into the proper format (WideBand (WB) codec) by the sender.
3) The transport module produces RTP/UDP/IP packets to be transmitted over the packet network.
4) The source call generator sends the IP packets to NIST Net emulator through IP address 1.
5) NIST Net applies the selected network conditions (delay, jitter, and packet loss).
6) NIST Net sends the result of the emulation to the receiver through IP address 2.
7) The receiver obtains the packet load.
8) The WideBand (WB) information is decoded and recorded into WAV format.

7.5.2 Simulating network conditions

The network conditions that the IP network emulator can emulate are the following:

- packet delay: fixed and variation (jitter);
- packet reordering;
- packet loss;
- packet duplication; and
- bandwidth limitation.

NIST Net uses a set of statistical routines to generate these parameters. The routines used need not be realistic models of the actual internal mechanisms of networks and routers, but should be computationally simple, and yet adaptable enough to imitate a wide range of network behaviours. That generation of values needed for delays and the rest of effects must be fast enough to keep up with packet receive rates, and any kernel based calculations must be done using only integral or fixed-point arithmetic. The routines should be parameterized in terms of quantities (such as mean, standard deviation, correlation, etc.) that are easily understood and easily derived from sources such as traffic traces, so that NIST Net can be made to imitate the behaviour demonstrated by those traces. Next, the techniques used to emulate the different network effects are described.

7.5.2.1 Delay and jitter emulation

Ordinary packet delay may be fixed or random, with the shape of the random distribution (run-time) settable. By default, an empirically-derived "heavy-tail" distribution is used. The mean, standard deviation, and linear correlation of packet-to-packet delays are all settable parameters. By appropriate setting of these parameters, packet reordering can be maximized, minimized or even eliminated. Packet delays may be specified with microsecond level precision; the actual delays imposed are limited by the granularity of the clock used.

The statistical distribution of the delay variation (jitter) is a matter of special interest. NIST Net by default implements the Pareto distribution. This distribution offers a good reproduction of the situation in real wide area networks. The probability for delayed packets is higher than the probability for packets coming earlier than the fixed delay time. A typical distribution of the delay jitter is shown in figure 12. This example is based on the evaluation of about 2 000 RTP packets under condition of 100 ms fixed delay and 20 ms delay jitter.

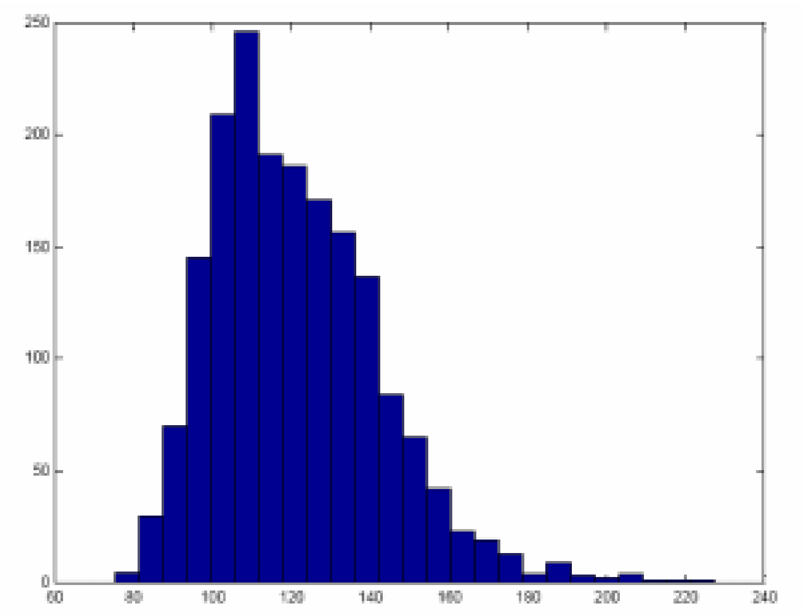


Figure 12: Example for the distribution of jitter generated by NIST Net

It is important to take into account some issues regarding NIST Net performance. If the jitter length set in the NIST Net is higher than the packet length, the amount of packets which arrive in the wrong order is strongly increased. Setups using small packet lengths are therefore disadvantaged. The packet ordering "rate" can not be set separately and precisely in the NIST Net. Also, band-limited transmission, such as xDSL, disturbs the RTP data stream. The control messages send out by the SIP and the H.323 protocols are that long, that they interrupt the RTP packet stream. DSL routers have additional buffers to compensate for this. The additional delay of the buffer is slowly removed (depending on the packet length used). This behaviour can originally not be emulated by NIST Net, but by the use of additional scripts. In order to consider these issues, a monitoring tool (e.g. Ethereal) should be used. Ethereal is a widely used open source packet analyzer, released under the GNU General Public Licence. It captures network packets and displays the packet data with detailed information. Ethereal currently runs on most UNIX and various Windows platforms, providing the possibility to capture and display live packet data from a network interface, open and save packet data captured, import and export packet data from and to a lot of other capture programs, filter and search packets on many criteria, create different statistics, etc.

For the purpose of ETSI STF 294, the default, empirically-derived "heavy-tail" distribution has been used.

7.5.2.2 Packet loss emulation

Packet loss distribution can follow very different laws depending on the cause for this impairment or on the type of network considered. NIST Net is able to implement most of them, either by using configuration or by replaying a script or a scenario recorded in real network conditions with a tool like Ethereal (as mentioned above for jitter distribution).

The most common distribution law is the random one. A shaped random number generator is used by NIST Net to drop packets in order to achieve a specific packet loss rate. If, for example, the packet loss rate is configured to 1 %, the NIST Net device would need to drop one packet out of 100. Because of the underlying random number generator it may occur that out of 100 packets 0 or 2 packets will be dropped. The configured packet loss rate will actually be achieved just after a large number of packets (> 1 000) with a certain maturity.

Other distribution laws are reflecting more bursty behaviours, where the events of packet loss are correlated and can result in a more important perception of cuts in the signal.

For the purpose of ETSI STF 294, random packet losses only have been used.

7.5.3 Network conditions

The network simulation database and the resulting recording of the speech sample database can be carried out using NIST Net running on a Linux-PC by varying the following:

- packet loss;
- delay; and
- jitter.

7.5.3.1 Typical values of network conditions

Table 3 shows the typical values for some of the fixed delay components.

Table 3: Typical values for fixed delay components

	Frame size (ms)	Look-ahead (ms)	Codec delay (ms) (see note 1)	Packetization delay (ms) (see note 2)	Serialization delay (ms) (see note 3)	De-jitter delay (ms)
G,722	0,125	1,5	1,625	0,125	negligible	20
G,722,1	20	20	40	20	negligible	20
G,722,2	20	5	25	20	negligible	20
Speex WB	20	14	34	20	negligible	20

NOTE 1: Codec delay = frame size + look-ahead.
NOTE 2: Packetization delay = frame size × frames/packet. Restriction < 30 ms.
NOTE 3: Serialization delay = frame size (bytes)/line speed. In most of scenarios it is considered negligible due to the high line speed and small cell size. At worst it does not exceed 40 ms.
NOTE 4: De-jitter delay = 1,5 × (sum of the variable delays). Typical 20 ms.

The other component that contributes to fixed end-to-end delay is the network delay, which will be modified to emulate different situations. It is formed by:

- **switching fixed delay**: although this delay is usually small, the architecture of the router or switch is the deciding factor. If a packet must be further buffered as a part of its processing, greater latency is incurred. Switching delays on today's high performance routers are negligible, typically in the order of 10 µs to 20 µs per packet; and

- **transmission delay:** it only becomes an important issue when the packets travel a great distance specifically, this delay is usually negligible if links are shorter than 1 000 km (see TR 101 329-6 [3]). Therefore this latency is function of the distance and the kind of the physical link, as can be found in G.114 [4].

Typical values for variable delay components.

- **queuing/buffering delay**: in most of scenarios it is considered negligible (not exceed 10 ms); and
- switching variable delay: 0 ms to 100 ms.

7.6 Network simulation database

The generally-accepted limit for good-quality voice connection delay is 200 ms one-way (or 250 ms as a limit). As delays rise over this figure, talkers and listeners become un-synchronized, and often they speak at the same time, or both wait for the other to speak. This condition is commonly called talker overlap. While the overall voice quality is acceptable, users may find the stilted nature of the conversation unacceptably annoying. Talker overlap may be observed on international telephone calls which travel over satellite connections (satellite delay is in the order of 500 ms, 250 ms up and 250 ms down). ITU recommends one-way speech delay < 150 ms (400 ms as an absolute limit). However, experience has shown that an end-to-end delay of 200 ms is still usually satisfactory for most users.

Typically, people can tolerate delays of no more than about 200 ms to 250 ms before the conversation gets annoying. Delays of 400 ms to 500 ms make conversations impractical.

Regarding information loss, ITU recommends packet loss < 3 % for audio communications, while jitter should not be more than 20 ms to 50 ms (< 1ms after de de-jitter buffering), as can be found in ITU-T Recommendation G.1010 [1].

Taking into account the typical values of the parameters and its requirements for audio communications, an example for conditions emulated with NIST Net is shown in table 4. It correspond to the choice made actually by ETSI STF 294.

Table 4: NIST Net simulation conditions

	End-to-end delay (ms)	Jitter (ms)	Packet loss (%)
1	0	0	0
2	150	10	1
3	400	20	3
Delay/Jitter distribution rule: "heavy-tail" Packet loss distribution rule: random			

It must be noted that for many potential database applications (like subjective listening tests) delay itself has no influence.

In order to simulate described network conditions, signals including speech plus background noise filtered out using Flt135 have been used by STF 294 due to the fact that this filter is probably the most typical one. Also 2 WideBand (WB) codecs (ITU-T Recommendation G.722 [6] and AMR WB or ITU-T Recommendation G.722.2 [9]) and network conditions shown in table 6 have been considered by STF 294.

8 Speech sample database description and subjective scores production

8.1 Database description

Output of the work described in the present document is the creation of a database containing the result of applying some typical and realistic transmission network scenarios and traffic patterns to a selected group of speech recordings. These speech samples are representative of different speech behaviours (language, speaker gender, speaker tone) and different WideBand (WB) terminal behaviours in the most typical noisy background ambiences.

Each of these conditions have been described in the previous clause, but to sum up the tables bellow collect all of this information. Table 5 enumerates conditions of speech samples with background noise before any terminal behaviour or network impairment, as actually implemented for ETSI STF 294.

Table 5: Speech sample database description - speech samples with background noise

Condition description	Number of conditions	Total
Languages	French Czech	2
Speakers	2 males 2 females	4
Noisy background	Cafeteria noise Office room noise Road noise Crossroads Car noise	5
Microphone-loudspeaker positions	Typical handset microphone position (with loudness ratings adjusted to 7 dB) Hands-free microphone position (with loudness ratings adjusted to 11 dB)	2
TOTAL		80 (2 × 4 × 5 × 2)

Each of the 80 samples in table 5 have been recorded in situations in table 6.

Table 6: Speech sample database description - noise reduction, coding and network transmission conditions

Condition description	Number of conditions			Total
Noise reduction (Flt 135 filter)	No noise estimation			9
	Noise estimation using VAD	Smooth noise reduction filter	Noise reduction level of 9 dB	
			Noise reduction level of 18 dB	
		Sharp noise reduction filter	Noise reduction level of 9 dB	
			Noise reduction level of 18 dB	
	Continuous noise estimation (no VAD)	Smooth noise reduction filter	Noise reduction level of 9 dB	
			Noise reduction level of 18 dB	
		Sharp noise reduction filter	Noise reduction level of 9 dB	
			Noise reduction level of 18 dB	
Coding	G,722 AMR-WB			2
Network impairments	Delay 0ms; Jitter 0ms; Loss 0 % (No impairments) Delay 150 ms; Jitter 10 ms; Loss 1 % Delay 400 ms; Jitter 20 ms; Loss 3 %			3
TOTAL				54 (9 × 2 × 3)

Total number of samples for STF 294: 54 × 80 = 4 320.

8.2 Subjective scores collection

8.2.1 Requirements

Subjective tests should provide a database with a wide range of possible impairments, both for objective algorithm training and validation.

8.2.2 Expert selection of samples for subjective testing

The original sample database recorded by STF 294 contains **4 320 conditions** corresponding to all the possible combinations of the factors as described in the Clause 8.1.

Being not feasible to run subjective testing on 4 320 speech samples, additional rules for qualified database size reduction have been identified. The following considerations have been made and confirmed by expert listening test:

- Since speakers for a same gender are not expected to generate important quality variations, only one speaker per gender is considered.
- In addition, the codec is expected to be a weak factor of quality variations, **compared with other factors** Therefore, the two codecs are distributed on different noises in the following way: AMR WB for noises perceived in mobile use situations such as roads, crossroads and car, and ITU-T Recommendation G.722 [6] for noises that can be perceived in fixed phone use situations such as office and cafeteria.
- In addition, it appeared that for road and cafeteria noises, speech signals are either inaudible or unintelligible with hands-free recording. Therefore, these conditions are removed.

Therefore, with these three considerations, the initial number of conditions (4 320) becomes 864, i.e. **432 conditions per language**. This is the amount of conditions that has been actually tested for the purpose of STF 294.

8.2.3 Methodology

This clause gives an overview of the subjective methodology. More detailed information on the practical procedures implemented by the Laboratories is available in annex B.

The methodology used for the subjective test is similar to the recommendation ITU-T Recommendation P.835 [25], since this methodology is really appropriate to noisy speech. It has been shown that for noisy (and especially de-noised) samples, only overall question does not lead to unified scoring [26].

Each trial contains three presentations of one sample, each presentation is followed by a silent voting period of 4 s. Each sample is 4 s in duration including about 1 s of background noise alone, 2 s of speech + noise, 1 s of background noise. Therefore, the total duration of each trial is 24 s.

For the two first presentations, listeners rate either the signal or the background depending on the rating scale order specified for that trial. For the signal, subjects are instructed to attend *only* to the *speech signal* and rate the speech on the five-category distortion scale shown in figure 13. For the background, subjects are instructed to attend *only* to the *background* and rate the background on the five-category intrusiveness scale shown in figure 14. For the third presentation in each trial, subjects are instructed to listen to the speech + background and rate it on the five-category overall quality scale shown in figure 15, the Mean Opinion Score (MOS) used with the ACR.

The exact instructions used by French and Czech laboratories during subjective testing are given in annex C.

To control for the effects of rating scale order, the order of the rating scales is balanced across the experiment, i.e. scale order should be "Signal, Background, Overall Effect" for half of the trials, and "Background, Signal, Overall Effect" for the other half. Furthermore, rating scale order is counter-balanced across listening panels.

```
Session 1       Block 1         Trial 1

Attending ONLY to the SPEECH SIGNAL, select the category
which best describes the sample you just heard.


             the SPEECH SIGNAL in this sample was

              5 – NOT DISTORTED

              4 – SLIGHTLY DISTORTED

              3 - SOMEWHAT DISTORTED

              2 – FAIRLY DISTORTED

              1 – VERY DISTORTED
```

Figure 13: Speech signal rating scale

```
Session 1       Block 1         Tr  ial 1

Attending ONLY to the BACKGROUND , select the category
which best describes the sample you just heard.


              the BACKGROUND  in this sample was

               5 – NOT PERCEPTIBLE

               4 – PERCEPTIBLE BUT NOT ANNOYING

               3 – SLIGHTLY ANNOYING

               2 – ANNOYING

               1 – VERY ANNOYING
```

Figure 14: Background noise rating scale (adopted from ITU-T Recommendation P.831 [27])

```
Select the category which best describes the sample you
just heard for purposes of everyday speech communication.

             the OVERALL SPEECH SAMPLE was

                   5 – EXCELLENT

                   4 – GOOD

                   3 - FAIR

                   2 – POOR

                   1 – BAD
```

Figure 15: Overall quality rating scale (same as the MOS rating scale) used in the ACR procedure (see ITU-T Recommendation P.800 [28])

The exact questionnaires used by French and Czech laboratories during subjective testing are given in annex B.

For the purpose of STF 294, the testing deployed 48 naive listeners (24 listeners per language). High-quality headphones with low-noise digital playback system have been used.

The resulting MOS scores must then be split into two parts: one for algorithm training (about 70 %) and one for the algorithm validation (30 %). The distribution between training and validation part have been tested by STF 294 to verify the equality of MOS and standard deviation distributions with respect to the entire database, according to a Kolmogorov-Smirnov test. The proportionality of occurrence of each parameter is tested to be as close as possible to 70 : 30 ratio. This ratio has been applied by STF 294.

8.2.4 STF 294 test results

The detailed results of the subjective tests performed for the purpose of STF 294 are shown in annex I as described in clause 8.2.4 (the results for the second group are for obvious reasons blinded). The proportionality of occurrence of each parameter is shown in table 7.

Table 7: Parameter value occurrence after database division into training and validation parts

Parameters	Training	Validation	Total	Training (%)	Validation (%)	Total (%)
Noise 1	39	15	54	72,2	27,8	100
Noise 2	36	18	54	66,7	33,3	100
Noise 3	35	19	54	64,8	35,2	100
Noise 4	34	20	54	63,0	37,0	100
Noise 5	43	11	54	79,6	20,4	100
Noise 6	39	15	54	72,2	27,8	100
Noise 7	36	18	54	66,7	33,3	100
Noise 8	40	14	54	74,1	25,9	100
Female	145	71	216	67,1	32,9	100
Male	157	59	216	72,7	27,3	100
Network1	98	46	144	68,1	31,9	100
Network2	103	41	144	71,5	28,5	100
Network3	101	43	144	70,1	29,9	100
NSA 1	34	14	48	70,8	29,2	100
NSA 2	31	17	48	64,6	35,4	100
NSA 3	35	13	48	72,9	27,1	100
NSA 4	33	15	48	68,8	31,3	100
NSA 5	30	18	48	62,5	37,5	100
NSA 6	40	8	48	83,3	16,7	100
NSA 7	36	12	48	75,0	25,0	100
NSA 8	26	22	48	54,2	45,8	100
NSA 9	37	11	48	77,1	22,9	100

Figure 16 shows the overall MOS and the associated 95 % CI obtained for each noise and for each language. For each noise environment, the handset and hands-free recording has been analyzed separately. It can be stated that the results match well between the two tested languages. The maximum differences occur for both Office noise conditions and do not exceed 0,8 MOS. The subjective scores between app. 1,5 to 4,3 cover the typical quality range.

While the relative performance of all tested coders and noise-suppressing techniques remains broadly similar, it is obvious that absolute values of MOS coming from different (distant) laboratories like Czech and French can not be compared directly even for identical (except of language) input speech material since the Mean Opinion Score term intrinsically contains many cultural and other aspects that are changing with time and location, affecting humans" quality demands.

Figure 17 depicts the overall MOS scores and associated 95 % CI obtained for each network condition, for each language separately. The Network conditions are as follows:

- NETWORK I: delay jitter 0ms, packet loss 0 %;
- NETWORK II: delay jitter 10ms, packet loss 1 %;
- NETWORK III: delay jitter 20ms, packet loss 3 %.

Figure 18 shows the NSA influence to the resulting scores. It seems that subjects' judgements have not been influenced by different NSA types, considering also large variability of noise and network conditions.

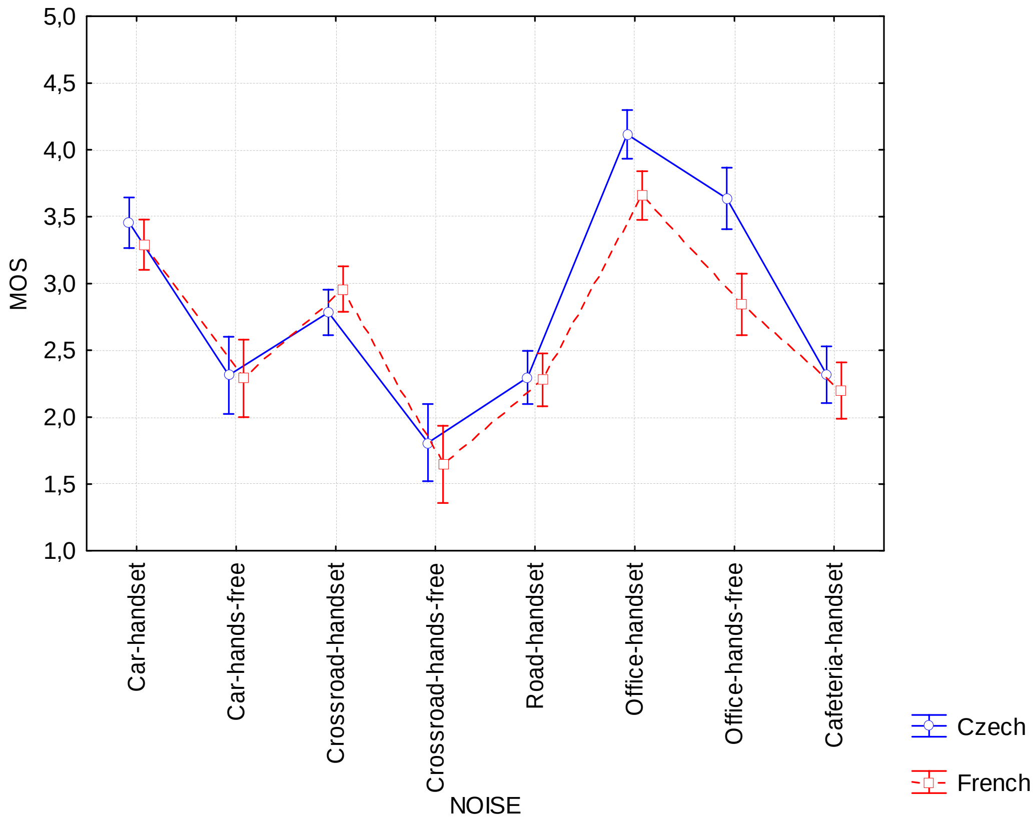
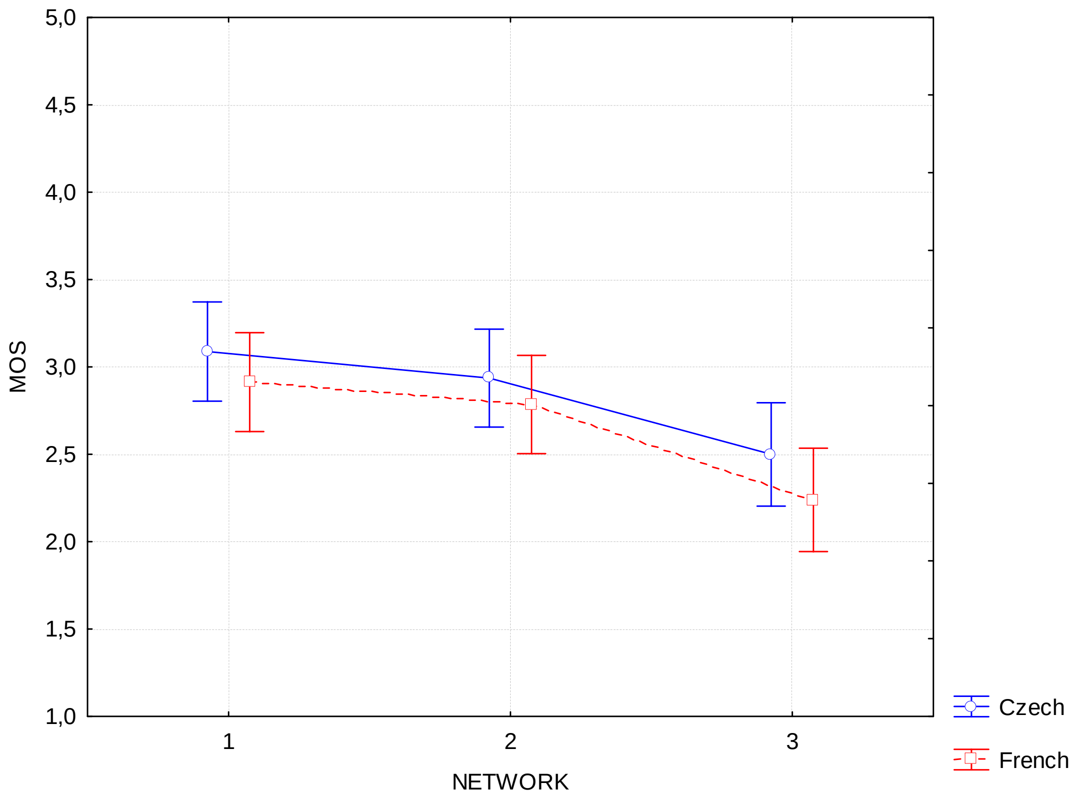


Figure 16: MOS distribution and associated 95 % CI per noise condition for Czech and French data



NOTE: NETWORK I: delay jitter 0ms. packet loss 0 %. NETWORK II: delay jitter 10ms. packet loss 1 %. NETWORK III: delay jitter 20ms. packet loss 3 %.

Figure 17: MOS distribution and associated 95 % CI per network condition for Czech and French data

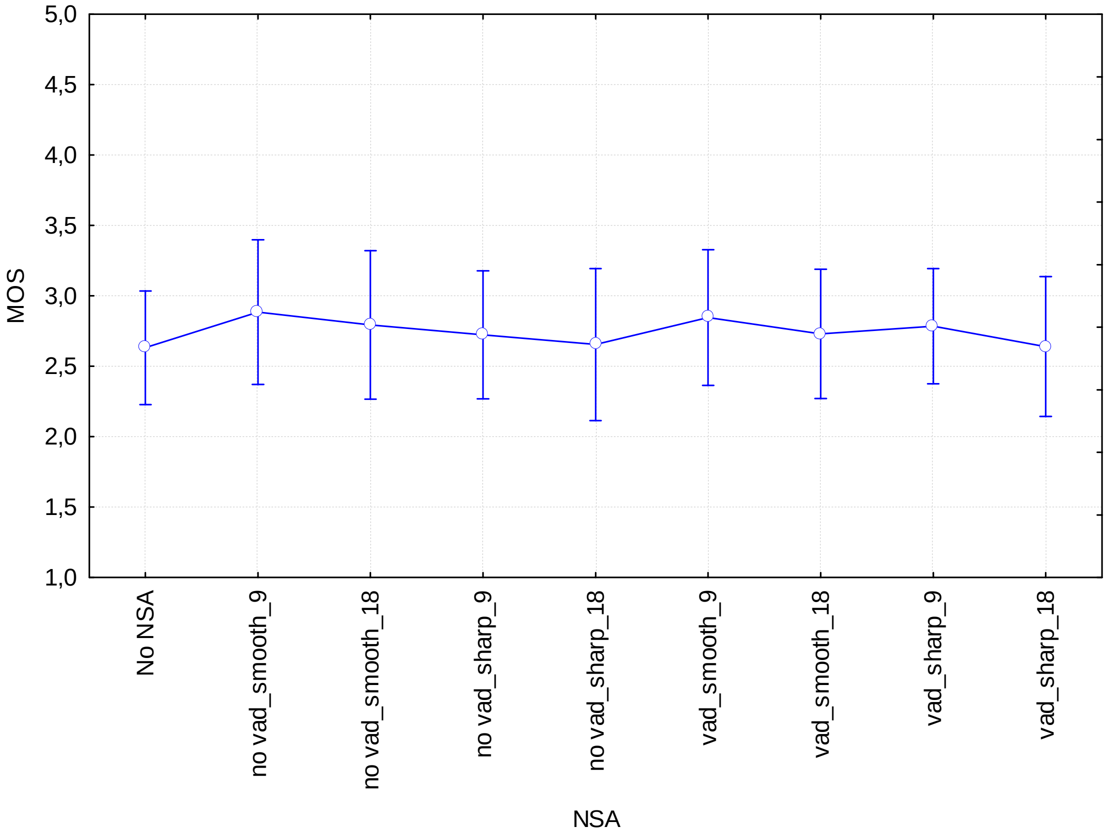


Figure 18: MOS distribution and associated 95 % CI per NSA condition

9 Application of the material produced

As mentioned in the scope. the setup and network simulation database as described in the document are applicable for different means. This clause tries to give some details on these applications.

9.1 Contribution of terminals and networks

It has been explained in clause 7.1 the database is containing speech samples collected at different points on the transmission process. At each of these points is being progressively added a new degradation or impairment. The recordings are:

1) noise-free speech samples;
2) speech samples with background noise conditions;
3) speech samples with noise and terminal conditions; and
4) speech samples with noise. terminal conditions. WideBand (WB) coding and network impairments.

Having registered in the database speech samples at these different points will allow anyone to listen and check how terminal processing. WideBand (WB) codec and network conditions are independently contributing to degrade the speech signal, by comparing the results before and after applying the correspondent condition. It can be used for a first quick evaluation of terminal processing options, extreme network condition, WideBand (WB) codec to be applied for a concrete usage (for instance, to decide whether to use sharp or smooth reduction filter in cafeteria noise situations. etc.).

9.2 Development of objective methods

As a typical example of such an application, the setup described above in clause 7 has been designed for the needs of ETSI Specialized Task Force (STF) 294 entitled: "Improving the quality of eEurope WideBand (WB) speech applications by developing a standardized performance testing and evaluation methodology for background noise transmission".

All the material produced, and in particular the speech sample database, has then been used in the frame of this STF in order to develop a new standard model for the evaluation of background noise transmission performance, as described in part 3 of the present document. A subset of this material has first been selected for subjective testing, whose results have then be used as a reference on which the model has then been trained and validated. A further subset is kept for a final independent validation of the model.

It must be mentioned that the subjective test, as well as the model developed under STF 294, are meant for a listening context. That means that they do not take into account the dimension of delay in the simulation.

Annex A (informative): Detailed STF 294 subjective test results

Table A.1 contains the sample conditions and related Speech MOS. Noise MOS and Global MOS for two tested languages. Also standard deviations for all MOS scores are given. The results for validation purposes are blinded.

Table A.1: Subjective experiment results - training part
(Subset: v - validation. t - training. Recording: hs - handset. hf - hands-free.Speaker: f - female. m - male)

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global
1	v	Car	hs	f	AMR _NI	no NSA	no NSA	no NSA												
2	v	Car	hs	f	AMR _NII	no NSA	no NSA	no NSA												
3	t	Car	hs	f	AMR _NIII	no NSA	no NSA	no NSA	3,08	2,79	2,83	0,72	0,66	0,56	3,63	3,13	3,08	0,88	0,61	0,72
4	v	Car	hs	f	AMR _NI	no	Smooth	9												
5	v	Car	hs	f	AMR _NII	no	Smooth	9												
6	v	Car	hs	f	AMR _NIII	no	Smooth	9												
7	t	Car	hs	f	AMR _NI	no	Smooth	18	4,00	3,83	3,92	0,66	0,48	0,50	4,21	3,71	3,63	0,93	0,62	0,71
8	t	Car	hs	f	AMR _NII	no	Smooth	18	3,29	3,54	3,25	0,81	0,51	0,68	4,21	3,71	3,58	0,83	0,86	0,78
9	v	Car	hs	f	AMR _NIII	no	Smooth	18												
10	v	Car	hs	f	AMR _NI	no	Sharp	9												
11	v	Car	hs	f	AMR _NII	no	Sharp	9												
12	t	Car	hs	f	AMR _NIII	no	Sharp	9	2,67	2,96	2,54	0,82	0,81	0,59	3,54	3,08	3,04	1,25	1,06	0,86
13	t	Car	hs	f	AMR _NI	no	Sharp	18	3,58	3,92	3,63	0,78	0,58	0,49	3,79	3,83	3,42	0,83	0,64	0,72
14	t	Car	hs	f	AMR _NII	no	Sharp	18	3,13	3,75	3,21	0,68	0,53	0,59	3,71	3,97	3,54	0,95	0,81	0,78
15	t	Car	hs	f	AMR _NIII	no	Sharp	18	2,83	3,58	2,88	0,82	0,50	0,74	2,42	3,29	2,38	0,88	0,95	0,71
16	t	Car	hs	f	AMR _NI	yes	Smooth	9	4,46	3,21	3,96	0,59	0,72	0,75	4,54	3,67	3,92	0,59	0,82	0,72
17	t	Car	hs	f	AMR _NII	yes	Smooth	9	3,75	3,29	3,67	0,44	0,62	0,56	3,83	3,58	3,33	0,96	0,78	0,92
18	t	Car	hs	f	AMR _NIII	yes	Smooth	9	2,42	3,25	2,58	0,72	0,53	0,65	3,29	3,04	2,96	0,86	0,69	0,55
19	t	Car	hs	f	AMR _NI	yes	Smooth	18	3,79	4,00	3,79	0,66	0,66	0,41	4,08	3,42	3,46	0,58	0,58	0,59

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
20	t	Car	hs	f	AMR _NII	yes	Smooth	18	2,92	3,63	3,17	0,72	0,65	0,56	3,88	3,63	3,45	0,99	0,71	0,65
21	t	Car	hs	f	AMR _NIII	yes	Smooth	18	2,75	3,25	2,54	0,85	0,85	0,72	2,63	3,67	2,58	0,97	0,70	0,97
22	v	Car	hs	f	AMR _NI	yes	Sharp	9												
23	t	Car	hs	f	AMR _NII	yes	Sharp	9	3,79	2,92	3,38	0,93	0,72	0,65	4,50	4,00	3,88	0,66	0,66	0,61
24	v	Car	hs	f	AMR _NIII	yes	Sharp	9												
25	t	Car	hs	f	AMR _NI	yes	Sharp	18	3,29	3,92	3,33	0,86	0,58	0,82	4,00	3,75	3,50	0,98	0,79	0,78
26	t	Car	hs	f	AMR _NII	yes	Sharp	18	3,58	3,79	3,54	0,83	0,51	0,66	3,63	4,08	3,63	0,97	0,65	0,71
27	t	Car	hs	f	AMR _NIII	yes	Sharp	18	3,04	3,79	3,04	0,86	0,41	0,62	2,83	3,88	2,63	1,20	0,80	1,01
28	t	Car	hf	f	AMR _NI	no NSA	no NSA	no NSA	3,54	1,50	2,17	0,88	0,66	0,87	3,79	2,25	2,54	1,10	0,79	0,72
29	t	Car	hf	f	AMR _NII	no NSA	no NSA	no NSA	3,21	1,58	2,25	0,93	0,65	0,74	3,71	2,08	2,46	1,08	1,10	0,93
30	t	Car	hf	f	AMR _NIII	no NSA	no NSA	no NSA	2,88	1,67	1,92	0,95	0,56	0,72	2,96	2,17	2,21	1,12	0,82	0,59
31	v	Car	hf	f	AMR _NI	no	Smooth	9												
32	v	Car	hf	f	AMR _NII	no	Smooth	9												
33	t	Car	hf	f	AMR _NIII	no	Smooth	9	2,83	2,08	2,29	0,82	0,83	0,69	2,96	2,04	2,25	0,95	0,62	0,74
34	v	Car	hf	f	AMR _NI	no	Smooth	18												
35	t	Car	hf	f	AMR _NII	no	Smooth	18	3,00	2,50	2,75	0,88	0,83	0,74	2,54	2,58	2,33	0,66	0,78	0,48
36	t	Car	hf	f	AMR _NIII	no	Smooth	18	1,58	2,83	1,79	0,58	0,76	0,72	2,51	1,79	1,75	0,88	0,78	0,68
37	v	Car	hf	f	AMR _NI	no	Sharp	9												
38	t	Car	hf	f	AMR _NII	no	Sharp	9	3,79	1,79	2,67	0,66	0,59	0,64	3,38	2,17	2,29	0,82	0,56	0,55
39	v	Car	hf	f	AMR _NIII	no	Sharp	9												
40	t	Car	hf	f	AMR _NI	no	Sharp	18	2,83	2,42	2,38	0,64	0,72	0,49	2,42	2,17	2,08	0,83	0,64	0,65
41	t	Car	hf	f	AMR _NII	no	Sharp	18	2,83	2,46	2,42	0,76	0,72	0,65	2,38	2,29	2,13	0,92	0,91	0,54
42	v	Car	hf	f	AMR _NIII	no	Sharp	18												
43	v	Car	hf	f	AMR _NI	yes	Smooth	9												
44	t	Car	hf	f	AMR _NII	yes	Smooth	9	3,54	1,92	2,63	0,72	0,65	0,88	3,50	1,92	2,38	0,93	0,65	0,58
45	t	Car	hf	f	AMR _NIII	yes	Smooth	9	2,54	2,38	2,17	0,93	0,71	0,70	2,53	1,88	1,83	0,83	0,90	0,70
46	t	Car	hf	f	AMR _NI	yes	Smooth	18	3,08	2,04	2,46	0,78	0,81	0,83	3,25	1,71	2,13	0,99	0,69	0,68
47	v	Car	hf	f	AMR _NII	yes	Smooth	18												
48	v	Car	hf	f	AMR _NIII	yes	Smooth	18												
49	v	Car	hf	f	AMR _NI	yes	Sharp	9												
50	t	Car	hf	f	AMR _NII	yes	Sharp	9	3,04	1,92	2,42	0,62	0,65	0,50	3,17	2,13	2,25	1,09	0,85	0,79
51	t	Car	hf	f	AMR _NIII	yes	Sharp	9	2,25	1,75	1,88	0,61	0,61	0,54	2,75	1,92	2,13	1,19	0,65	0,61
52	v	Car	hf	f	AMR _NI	yes	Sharp	18												
53	t	Car	hf	f	AMR _NII	yes	Sharp	18	2,25	2,17	1,96	0,61	0,82	0,75	2,08	1,96	1,67	0,88	0,75	0,70
54	t	Car	hf	f	AMR _NIII	yes	Sharp	18	1,83	2,04	1,54	0,76	1,00	0,66	2,00	1,92	1,63	0,93	0,83	0,65
55	t	Car	hs	m	AMR _NI	no NSA	no NSA	no NSA	3,75	2,88	3,29	0,61	0,90	0,55	4,33	3,04	3,21	0,76	0,62	0,66
56	t	Car	hs	m	AMR _NII	no NSA	no NSA	no NSA	3,92	3,08	3,46	0,58	0,58	0,51	4,58	3,50	3,58	0,58	0,66	0,50
57	t	Car	hs	m	AMR _NIII	no NSA	no NSA	no NSA	3,92	3,00	3,29	0,65	0,59	0,55	3,46	3,00	2,79	0,98	0,98	0,78
58	t	Car	hs	m	AMR _NI	no	Smooth	9	4,54	4,00	4,42	0,59	0,42	0,50	4,79	3,17	3,58	0,41	0,64	0,58

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
59	t	Car	hs	m	AMR _NII	no	Smooth	9	4,17	3,88	4,17	0,76	0,61	0,64	4,21	3,50	3,38	0,98	0,66	0,88
60	t	Car	hs	m	AMR _NIII	no	Smooth	9	3,79	3,46	3,54	0,66	0,51	0,59	3,29	3,00	2,88	1,12	0,83	0,90
61	t	Car	hs	m	AMR _NI	no	Smooth	18	3,79	4,17	3,88	0,78	0,48	0,54	4,08	3,79	3,63	0,97	0,59	0,71
62	t	Car	hs	m	AMR _NII	no	Smooth	18	3,96	4,13	4,00	0,62	0,54	0,51	3,75	3,88	3,54	0,94	0,61	0,88
63	v	Car	hs	m	AMR _NIII	no	Smooth	18												
64	t	Car	hs	m	AMR _NI	no	Sharp	9	4,42	3,25	3,96	0,58	0,68	0,69	4,17	2,91	3,17	0,70	0,72	0,56
65	t	Car	hs	m	AMR _NII	no	Sharp	9	3,92	3,29	3,58	0,50	0,55	0,50	4,29	3,25	3,46	0,75	0,85	0,78
66	t	Car	hs	m	AMR _NIII	no	Sharp	9	2,33	3,04	2,46	1,01	0,62	0,72	3,58	2,79	2,92	0,97	1,02	0,83
67	v	Car	hs	m	AMR _NI	no	Sharp	18												
68	t	Car	hs	m	AMR _NII	no	Sharp	18	3,17	3,96	3,50	0,96	0,46	0,78	3,67	3,67	3,13	0,82	0,64	0,74
69	t	Car	hs	m	AMR _NIII	no	Sharp	18	3,17	3,96	3,38	0,82	0,55	0,65	2,82	3,29	2,79	0,70	0,75	0,66
70	t	Car	hs	m	AMR _NI	yes	Smooth	9	4,63	3,54	4,25	0,49	0,72	0,61	4,65	3,75	3,71	0,76	0,68	0,75
71	t	Car	hs	m	AMR _NII	yes	Smooth	9	3,92	3,25	3,67	0,58	0,53	0,48	4,54	3,54	3,54	0,78	0,72	0,59
72	t	Car	hs	m	AMR _NIII	yes	Smooth	9	3,83	3,58	3,75	0,70	0,50	0,61	3,13	3,33	2,58	1,08	0,82	0,88
73	t	Car	hs	m	AMR _NI	yes	Smooth	18	4,17	4,08	4,17	0,76	0,41	0,38	3,88	3,54	3,54	0,90	0,78	0,59
74	t	Car	hs	m	AMR _NII	yes	Smooth	18	3,00	3,88	3,17	0,98	0,61	0,70	3,42	3,83	3,33	1,10	0,56	0,76
75	t	Car	hs	m	AMR _NIII	yes	Smooth	18	3,54	3,46	3,42	1,02	0,59	0,65	2,96	3,38	2,67	1,08	0,77	0,87
76	t	Car	hs	m	AMR _NI	yes	Sharp	9	4,42	3,25	3,88	0,50	0,61	0,61	4,25	3,38	3,50	0,74	0,65	0,66
77	t	Car	hs	m	AMR _NII	yes	Sharp	9	4,21	3,46	3,67	0,66	0,51	0,64	4,21	3,58	3,63	0,83	0,83	0,65
78	v	Car	hs	m	AMR _NIII	yes	Sharp	9												
79	v	Car	hs	m	AMR _NI	yes	Sharp	18												
80	t	Car	hs	m	AMR _NII	yes	Sharp	18	4,21	4,08	4,08	0,72	0,41	0,65	3,38	3,63	3,25	0,88	0,88	0,68
81	v	Car	hs	m	AMR _NIII	yes	Sharp	18												
82	t	Car	hf	m	AMR _NI	no NSA	no NSA	no NSA	3,58	1,42	2,17	1,14	0,58	0,82	4,00	2,21	2,54	0,98	0,78	0,66
83	t	Car	hf	m	AMR _NII	no NSA	no NSA	no NSA	3,29	1,67	2,33	0,86	0,64	0,64	3,92	1,92	2,54	1,06	0,72	0,59
84	t	Car	hf	m	AMR _NIII	no NSA	no NSA	no NSA	2,29	1,50	1,67	0,86	0,59	0,56	3,13	1,83	2,46	0,90	0,64	0,59
85	t	Car	hf	m	AMR _NI	no	Smooth	9	3,96	2,54	2,92	0,62	0,66	0,65	4,04	2,54	2,92	0,95	0,59	0,65
86	t	Car	hf	m	AMR _NII	no	Smooth	9	3,96	2,38	2,88	0,69	0,71	0,68	3,54	2,50	2,75	0,88	0,72	0,53
87	t	Car	hf	m	AMR _NIII	no	Smooth	9	2,13	2,13	1,96	0,74	0,74	0,62	2,71	2,00	2,21	0,69	0,78	0,72
88	t	Car	hf	m	AMR _NI	no	Smooth	18	3,13	2,96	2,71	0,68	0,95	0,81	3,29	2,58	2,75	1,04	1,02	0,90
89	t	Car	hf	m	AMR _NII	no	Smooth	18	2,71	2,88	2,42	0,69	0,90	0,58	3,04	2,88	2,67	1,12	1,03	0,92
90	v	Car	hf	m	AMR _NIII	no	Smooth	18												
91	t	Car	hf	m	AMR _NI	no	Sharp	9	1,88	2,83	2,00	0,61	0,82	0,66	3,46	2,04	2,54	0,78	0,81	0,72
92	t	Car	hf	m	AMR _NII	no	Sharp	9	3,46	1,96	2,38	0,93	0,62	0,71	3,42	2,43	2,71	0,97	0,65	0,62
93	t	Car	hf	m	AMR _NIII	no	Sharp	9	3,33	2,00	2,63	0,87	0,59	0,77	2,75	1,83	1,92	0,99	0,82	0,65
94	v	Car	hf	m	AMR _NI	no	Sharp	18												
95	t	Car	hf	m	AMR _NII	no	Sharp	18	2,67	2,79	2,58	0,56	0,72	0,78	2,71	2,54	2,35	0,95	0,93	0,81
96	t	Car	hf	m	AMR _NIII	no	Sharp	18	1,17	2,83	1,33	0,48	0,64	0,64	2,04	2,46	1,96	0,81	0,93	0,69
97	t	Car	hf	m	AMR _NI	yes	Smooth	9	3,88	2,29	3,08	0,80	0,69	0,72	4,08	2,50	2,83	0,72	0,88	0,64

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
98	t	Car	hf	m	AMR _NII	yes	Smooth	9	3,21	2,17	2,54	0,78	0,64	0,88	3,46	2,29	2,46	1,14	0,81	0,66
99	v	Car	hf	m	AMR _NIII	yes	Smooth	9												
100	t	Car	hf	m	AMR _NI	yes	Smooth	18	3,00	2,04	2,33	0,88	0,75	0,76	2,96	2,38	2,38	0,81	0,82	0,58
101	v	Car	hf	m	AMR _NII	yes	Smooth	18												
102	v	Car	hf	m	AMR _NIII	yes	Smooth	18												
103	v	Car	hf	m	AMR _NI	yes	Sharp	9												
104	v	Car	hf	m	AMR _NII	yes	Sharp	9												
105	t	Car	hf	m	AMR _NIII	yes	Sharp	9	3,46	2,08	2,46	0,78	0,58	0,66	2,79	2,50	2,38	1,02	0,66	0,71
106	t	Car	hf	m	AMR _NI	yes	Sharp	18	2,42	2,38	1,75	1,02	0,97	0,79	2,58	2,42	2,23	1,25	0,72	0,78
107	t	Car	hf	m	AMR _NII	yes	Sharp	18	2,92	2,04	2,00	0,72	0,91	0,78	2,42	2,46	1,96	1,18	0,72	0,75
108	t	Car	hf	m	AMR _NIII	yes	Sharp	18	2,88	2,04	2,33	0,90	0,62	0,56	2,17	2,21	2,04	1,01	0,93	0,62
109	t	Crossroads	hs	f	AMR _NI	no NSA	no NSA	no NSA	3,00	1,92	2,46	0,78	0,65	0,59	4,38	3,29	3,42	1,01	0,69	0,65
110	t	Crossroads	hs	f	AMR _NII	no NSA	no NSA	no NSA	3,08	2,17	2,63	0,78	0,64	0,58	3,88	2,92	3,04	1,03	0,78	0,62
111	t	Crossroads	hs	f	AMR _NIII	no NSA	no NSA	no NSA	2,21	1,88	1,88	0,78	0,61	0,61	3,29	3,00	2,67	0,95	0,59	0,64
112	t	Crossroads	hs	f	AMR _NI	no	Smooth	9	3,63	2,71	3,04	1,01	0,69	0,69	4,36	3,25	3,29	0,64	0,74	0,81
113	t	Crossroads	hs	f	AMR _NII	no	Smooth	9	2,92	2,04	2,54	0,78	0,69	0,51	4,48	3,33	3,71	0,77	0,76	0,75
114	v	Crossroads	hs	f	AMR _NIII	no	Smooth	9												
115	t	Crossroads	hs	f	AMR _NI	no	Smooth	18	2,75	3,42	2,96	0,74	0,65	0,46	4,04	3,46	3,67	0,81	0,51	0,64
116	v	Crossroads	hs	f	AMR _NII	no	Smooth	18												
117	t	Crossroads	hs	f	AMR _NIII	no	Smooth	18	1,96	3,25	2,54	0,81	0,68	0,66	3,71	3,58	3,08	1,12	0,72	0,83
118	t	Crossroads	hs	f	AMR _NI	no	Sharp	9	3,54	2,29	2,67	0,93	0,46	0,64	4,21	3,00	3,33	0,83	0,66	0,76
119	v	Crossroads	hs	f	AMR _NII	no	Sharp	9												
120	t	Crossroads	hs	f	AMR _NIII	no	Sharp	9	2,00	1,96	1,92	0,72	0,55	0,41	2,88	2,42	2,25	1,08	0,72	0,74
121	t	Crossroads	hs	f	AMR _NI	no	Sharp	18	2,71	3,08	2,79	0,75	0,65	0,66	4,08	3,29	3,38	0,78	0,62	0,71
122	v	Crossroads	hs	f	AMR _NII	no	Sharp	18												
123	t	Crossroads	hs	f	AMR _NIII	no	Sharp	18	2,00	2,88	2,04	0,78	0,61	0,55	2,50	3,52	2,50	0,83	0,77	0,66
124	t	Crossroads	hs	f	AMR _NI	yes	Smooth	9	3,79	2,46	3,17	0,78	0,51	0,56	4,50	3,46	3,50	0,66	0,59	0,59
125	t	Crossroads	hs	f	AMR _NII	yes	Smooth	9	2,54	2,17	2,38	0,59	0,64	0,58	4,38	3,46	3,50	0,77	0,66	0,51
126	t	Crossroads	hs	f	AMR _NIII	yes	Smooth	9	3,50	2,46	2,96	0,72	0,59	0,36	3,58	3,25	2,88	1,21	0,85	0,80
127	t	Crossroads	hs	f	AMR _NI	yes	Smooth	18	2,96	3,50	3,04	0,69	0,51	0,46	4,08	3,67	3,46	0,97	0,82	0,83
128	t	Crossroads	hs	f	AMR _NII	yes	Smooth	18	1,88	2,88	1,96	0,74	0,80	0,69	3,54	3,52	3,46	1,14	0,77	0,78
129	t	Crossroads	hs	f	AMR _NIII	yes	Smooth	18	1,58	2,79	1,71	0,72	0,93	0,55	3,13	3,57	3,00	1,03	0,71	0,78
130	t	Crossroads	hs	f	AMR _NI	yes	Sharp	9	3,50	2,42	2,92	0,88	0,65	0,65	4,46	3,33	3,38	0,66	0,70	0,58
131	t	Crossroads	hs	f	AMR _NII	yes	Sharp	9	3,25	2,08	2,58	0,79	0,65	0,58	4,08	3,54	3,50	1,02	0,72	0,93
132	v	Crossroads	hs	f	AMR _NIII	yes	Sharp	9												
133	t	Crossroads	hs	f	AMR _NI	yes	Sharp	18	2,00	3,38	2,17	0,72	0,58	0,64	3,63	4,04	3,71	1,13	0,75	0,86
134	v	Crossroads	hs	f	AMR _NII	yes	Sharp	18												
135	t	Crossroads	hs	f	AMR _NIII	yes	Sharp	18	2,00	2,67	1,88	0,88	0,87	0,74	2,96	3,46	2,83	0,91	0,98	0,56
136	v	Crossroads	hf	f	AMR _NI	no NSA	no NSA	no NSA												

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
137	t	Crossroads	hf	f	AMR _NII	no NSA	no NSA	no NSA	2,88	1,58	1,96	0,90	0,72	0,75	2,17	1,63	1,50	1,24	0,92	0,78
138	t	Crossroads	hf	f	AMR _NIII	no NSA	no NSA	no NSA	1,79	1,29	1,33	0,88	0,46	0,56	1,92	1,58	1,29	1,02	0,93	0,55
139	t	Crossroads	hf	f	AMR _NI	no	Smooth	9	3,17	2,13	2,33	0,76	0,74	0,76	3,21	1,38	1,73	1,47	0,65	0,74
140	t	Crossroads	hf	f	AMR _NII	no	Smooth	9	2,88	2,04	2,29	0,80	0,69	0,81	2,74	1,67	1,83	1,11	1,01	0,87
141	v	Crossroads	hf	f	AMR _NIII	no	Smooth	9												
142	t	Crossroads	hf	f	AMR _NI	no	Smooth	18	2,88	2,88	2,54	0,80	0,90	0,72	2,25	1,71	1,79	1,19	0,86	0,83
143	v	Crossroads	hf	f	AMR _NII	no	Smooth	18												
144	t	Crossroads	hf	f	AMR _NIII	no	Smooth	18	2,75	2,42	2,46	1,07	0,72	0,83	1,79	1,96	1,46	0,93	1,43	0,72
145	v	Crossroads	hf	f	AMR _NI	no	Sharp	9												
146	t	Crossroads	hf	f	AMR _NII	no	Sharp	9	2,92	2,08	2,42	0,83	0,72	0,88	2,71	1,33	1,54	1,33	0,56	0,59
147	t	Crossroads	hf	f	AMR _NIII	no	Sharp	9	2,21	1,54	1,71	0,59	0,59	0,62	2,29	1,21	1,54	1,20	0,59	0,66
148	t	Crossroads	hf	f	AMR _NI	no	Sharp	18	2,46	2,83	2,46	0,88	0,76	0,78	1,75	1,88	1,71	0,79	0,95	0,91
149	v	Crossroads	hf	f	AMR _NII	no	Sharp	18												
150	v	Crossroads	hf	f	AMR _NIII	no	Sharp	18												
151	t	Crossroads	hf	f	AMR _NI	yes	Smooth	9	3,04	2,13	2,21	1,04	0,68	0,83	2,96	1,54	1,71	1,37	0,66	0,81
152	t	Crossroads	hf	f	AMR _NII	yes	Smooth	9	2,67	1,79	2,08	0,92	0,51	0,72	2,79	1,38	1,58	1,32	0,58	0,72
153	t	Crossroads	hf	f	AMR _NIII	yes	Smooth	9	1,96	1,88	1,63	0,81	0,80	0,77	1,58	1,29	1,38	0,97	0,55	0,58
154	t	Crossroads	hf	f	AMR _NI	yes	Smooth	18	3,04	1,67	2,13	1,00	0,82	0,74	2,42	1,42	1,54	1,10	0,58	0,59
155	t	Crossroads	hf	f	AMR _NII	yes	Smooth	18	2,00	2,00	1,50	0,93	0,93	0,83	2,00	1,71	1,63	0,93	0,75	0,65
156	t	Crossroads	hf	f	AMR _NIII	yes	Smooth	18	3,08	1,79	2,17	0,88	0,83	0,64	1,83	1,25	1,29	1,09	0,44	0,55
157	v	Crossroads	hf	f	AMR _NI	yes	Sharp	9												
158	v	Crossroads	hf	f	AMR _NII	yes	Sharp	9												
159	t	Crossroads	hf	f	AMR _NIII	yes	Sharp	9	1,75	1,50	1,33	0,79	0,66	0,48	1,96	1,33	1,46	1,00	0,56	0,72
160	t	Crossroads	hf	f	AMR _NI	yes	Sharp	18	1,63	2,25	1,50	0,82	1,15	0,83	1,88	1,63	1,54	1,03	0,71	0,78
161	v	Crossroads	hf	f	AMR _NII	yes	Sharp	18												
162	t	Crossroads	hf	f	AMR _NIII	yes	Sharp	18	1,96	1,88	1,75	0,81	0,74	0,74	1,38	1,54	1,13	0,71	0,93	0,45
163	v	Crossroads	hs	m	AMR _NI	no NSA	no NSA	no NSA												
164	v	Crossroads	hs	m	AMR _NII	no NSA	no NSA	no NSA												
165	v	Crossroads	hs	m	AMR _NIII	no NSA	no NSA	no NSA												
166	t	Crossroads	hs	m	AMR _NI	no	Smooth	9	4,04	3,00	3,67	0,81	0,66	0,82	4,13	2,83	3,00	1,08	0,92	0,78
167	t	Crossroads	hs	m	AMR _NII	no	Smooth	9	3,21	2,83	2,79	0,88	0,56	0,59	3,88	2,83	2,71	1,03	1,05	0,91
168	t	Crossroads	hs	m	AMR _NIII	no	Smooth	9	3,42	2,75	3,08	0,97	0,68	0,58	2,96	2,42	2,29	1,27	0,88	0,91
169	t	Crossroads	hs	m	AMR _NI	no	Smooth	18	2,83	3,92	3,04	0,82	0,41	0,81	3,08	2,92	2,75	1,06	1,18	1,11
170	t	Crossroads	hs	m	AMR _NII	no	Smooth	18	3,50	3,75	3,58	0,88	0,44	0,72	3,04	2,96	2,75	1,04	1,08	1,11
171	v	Crossroads	hs	m	AMR _NIII	no	Smooth	18												
172	v	Crossroads	hs	m	AMR _NI	no	Sharp	9												
173	t	Crossroads	hs	m	AMR _NII	no	Sharp	9	3,67	3,13	3,21	0,56	0,61	0,51	4,13	2,58	3,00	0,68	0,88	0,66
174	t	Crossroads	hs	m	AMR _NIII	no	Sharp	9	2,42	2,38	2,00	0,93	0,58	0,66	2,75	2,08	2,00	1,07	0,93	0,98
175	t	Crossroads	hs	m	AMR _NI	no	Sharp	18	2,88	3,46	3,08	0,99	0,51	0,72	3,21	3,17	2,88	1,06	1,05	0,85

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
176	t	Crossroads	hs	m	AMR _NII	no	Sharp	18	3,17	3,29	3,08	0,87	0,62	0,72	3,00	3,50	3,13	0,78	0,88	0,74
177	v	Crossroads	hs	m	AMR _NIII	no	Sharp	18												
178	t	Crossroads	hs	m	AMR _NI	yes	Smooth	9	4,29	3,00	3,54	0,55	0,51	0,59	3,96	2,92	3,13	0,81	0,93	1,03
179	v	Crossroads	hs	m	AMR _NII	yes	Smooth	9												
180	t	Crossroads	hs	m	AMR _NIII	yes	Smooth	9	3,67	2,83	2,96	0,64	0,48	0,46	2,83	2,63	2,50	1,17	0,97	0,98
181	v	Crossroads	hs	m	AMR _NI	yes	Smooth	18												
182	v	Crossroads	hs	m	AMR _NII	yes	Smooth	18												
183	t	Crossroads	hs	m	AMR _NIII	yes	Smooth	18	1,92	2,88	1,96	0,93	0,45	0,69	3,25	3,00	2,79	1,15	1,29	1,22
184	v	Crossroads	hs	m	AMR _NI	yes	Sharp	9												
185	v	Crossroads	hs	m	AMR _NII	yes	Sharp	9												
186	v	Crossroads	hs	m	AMR _NIII	yes	Sharp	9												
187	v	Crossroads	hs	m	AMR _NI	yes	Sharp	18												
188	t	Crossroads	hs	m	AMR _NII	yes	Sharp	18	3,17	3,96	2,96	1,01	0,55	0,81	2,88	3,21	2,96	1,15	0,72	1,08
189	t	Crossroads	hs	m	AMR _NIII	yes	Sharp	18	3,46	3,88	3,33	0,98	0,45	0,76	3,25	3,46	2,67	1,15	0,93	0,87
190	t	Crossroads	hf	m	AMR _NI	no NSA	no NSA	no NSA	2,71	1,21	1,33	1,20	0,51	0,56	2,71	1,50	1,71	1,37	0,66	0,69
191	t	Crossroads	hf	m	AMR _NII	no NSA	no NSA	no NSA	2,50	1,13	1,29	0,93	0,34	0,46	2,93	1,86	1,92	1,34	1,08	0,97
192	v	Crossroads	hf	m	AMR _NIII	no NSA	no NSA	no NSA												
193	v	Crossroads	hf	m	AMR _NI	no	Smooth	9												
194	v	Crossroads	hf	m	AMR _NII	no	Smooth	9												
195	t	Crossroads	hf	m	AMR _NIII	no	Smooth	9	1,38	1,42	1,21	0,65	0,58	0,41	2,00	1,46	1,42	0,88	0,88	0,58
196	t	Crossroads	hf	m	AMR _NI	no	Smooth	18	1,88	2,25	1,92	0,80	0,79	0,58	2,38	1,88	1,92	1,13	1,12	0,72
197	t	Crossroads	hf	m	AMR _NII	no	Smooth	18	2,08	2,38	2,00	1,06	0,71	0,88	2,63	1,79	2,13	1,31	0,78	0,90
198	t	Crossroads	hf	m	AMR _NIII	no	Smooth	18	1,92	2,42	1,83	0,83	0,83	0,76	1,83	1,50	1,38	1,13	0,72	0,58
199	v	Crossroads	hf	m	AMR _NI	no	Sharp	9												
200	t	Crossroads	hf	m	AMR _NII	no	Sharp	9	2,50	1,33	1,75	1,06	0,56	0,74	2,92	1,58	1,75	1,32	0,78	0,74
201	v	Crossroads	hf	m	AMR _NIII	no	Sharp	9												
202	t	Crossroads	hf	m	AMR _NI	no	Sharp	18	1,71	2,25	1,54	0,75	0,85	0,78	1,96	1,71	1,79	1,04	0,81	0,78
203	v	Crossroads	hf	m	AMR _NII	no	Sharp	18												
204	v	Crossroads	hf	m	AMR _NIII	no	Sharp	18												
205	t	Crossroads	hf	m	AMR _NI	yes	Smooth	9	2,75	1,58	1,92	0,94	0,58	0,65	3,00	1,67	1,71	1,32	0,92	0,86
206	v	Crossroads	hf	m	AMR _NII	yes	Smooth	9												
207	t	Crossroads	hf	m	AMR _NIII	yes	Smooth	9	2,29	1,42	1,50	1,20	0,58	0,72	2,67	1,29	1,50	0,96	0,46	0,59
208	t	Crossroads	hf	m	AMR _NI	yes	Smooth	18	2,92	1,42	1,83	0,93	0,58	0,56	2,67	1,96	2,04	1,20	0,91	0,86
209	v	Crossroads	hf	m	AMR _NII	yes	Smooth	18												
210	t	Crossroads	hf	m	AMR _NIII	yes	Smooth	18	2,17	1,46	1,71	0,96	0,66	0,69	2,17	1,50	1,46	1,20	0,93	0,59
211	t	Crossroads	hf	m	AMR _NI	yes	Sharp	9	2,58	1,38	1,83	1,02	0,58	0,82	2,88	1,75	2,13	1,33	0,94	0,90
212	v	Crossroads	hf	m	AMR _NII	yes	Sharp	9												
213	v	Crossroads	hf	m	AMR _NIII	yes	Sharp	9												
214	t	Crossroads	hf	m	AMR _NI	yes	Sharp	18	1,25	2,00	1,29	0,61	1,10	0,62	1,92	2,13	1,55	1,02	1,12	0,71

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global
215	t	Crossroads	hf	m	AMR _NII	yes	Sharp	18	1,71	1,67	1,54	0,95	0,76	0,78	1,79	2,17	1,54	1,14	1,09	0,66
216	t	Crossroads	hf	m	AMR _NIII	yes	Sharp	18	1,79	1,79	1,54	0,66	0,66	0,59	1,92	1,67	1,54	0,88	0,70	0,59
217	t	Road	hs	f	AMR _NI	no NSA	no NSA	no NSA	2,50	1,67	1,92	0,83	0,64	0,50	3,50	2,17	2,39	1,02	0,64	0,49
218	v	Road	hs	f	AMR _NII	no NSA	no NSA	no NSA												
219	t	Road	hs	f	AMR _NIII	no NSA	no NSA	no NSA	1,67	1,50	1,50	0,64	0,51	0,59	3,29	2,21	2,17	1,00	0,98	0,64
220	t	Road	hs	f	AMR _NI	no	Smooth	9	3,04	1,96	2,58	0,81	0,69	0,72	4,25	2,14	2,71	0,79	0,80	0,62
221	t	Road	hs	f	AMR _NII	no	Smooth	9	2,83	1,71	2,25	0,70	0,62	0,61	3,83	2,04	2,63	0,82	0,86	0,82
222	t	Road	hs	f	AMR _NIII	no	Smooth	9	3,33	2,08	2,75	1,01	0,58	0,79	2,79	1,67	2,10	0,88	0,70	0,66
223	t	Road	hs	f	AMR _NI	no	Smooth	18	2,29	2,67	2,33	0,75	0,48	0,48	3,50	2,50	2,75	0,98	0,83	0,68
224	t	Road	hs	f	AMR _NII	no	Smooth	18	2,29	2,38	2,17	0,69	0,65	0,76	3,13	2,96	2,71	0,90	0,62	0,75
225	v	Road	hs	f	AMR _NIII	no	Smooth	18												
226	v	Road	hs	f	AMR _NI	no	Sharp	9												
227	t	Road	hs	f	AMR _NII	no	Sharp	9	2,33	1,71	1,96	0,64	0,55	0,62	3,33	2,75	2,58	0,82	0,85	0,72
228	t	Road	hs	f	AMR _NIII	no	Sharp	9	1,25	1,67	1,25	0,53	0,56	0,44	3,12	2,05	2,08	1,20	1,00	0,72
229	v	Road	hs	f	AMR _NI	no	Sharp	18												
230	t	Road	hs	f	AMR _NII	no	Sharp	18	2,08	2,63	2,08	0,83	0,58	0,72	2,67	2,88	2,42	0,87	0,68	0,58
231	t	Road	hs	f	AMR _NIII	no	Sharp	18	1,58	2,33	1,83	0,50	0,76	0,64	2,21	2,25	1,92	0,88	0,68	0,65
232	t	Road	hs	f	AMR _NI	yes	Smooth	9	3,25	1,75	2,42	0,79	0,44	0,78	4,00	2,29	2,88	0,88	0,55	0,61
233	t	Road	hs	f	AMR _NII	yes	Smooth	9	2,79	2,00	2,17	0,83	0,42	0,70	3,87	2,46	2,79	1,03	0,72	0,72
234	t	Road	hs	f	AMR _NIII	yes	Smooth	9	1,79	2,13	1,83	0,72	0,61	0,64	2,79	2,08	2,04	1,06	0,58	0,62
235	t	Road	hs	f	AMR _NI	yes	Smooth	18	2,33	2,79	2,29	0,70	0,66	0,55	3,46	2,83	2,71	0,98	0,76	0,69
236	t	Road	hs	f	AMR _NII	yes	Smooth	18	2,04	2,67	2,17	0,62	0,64	0,70	3,38	2,75	2,75	0,97	0,90	0,68
237	t	Road	hs	f	AMR _NIII	yes	Smooth	18	2,46	2,25	2,17	0,72	0,61	0,64	2,75	2,50	2,46	0,85	0,83	0,66
238	v	Road	hs	f	AMR _NI	yes	Sharp	9												
239	t	Road	hs	f	AMR _NII	yes	Sharp	9	2,50	1,71	1,83	0,83	0,62	0,56	3,25	2,67	2,83	0,79	0,82	0,64
240	t	Road	hs	f	AMR _NIII	yes	Sharp	9	1,71	1,63	1,50	0,69	0,49	0,51	3,00	2,17	2,08	0,88	0,70	0,72
241	v	Road	hs	f	AMR _NI	yes	Sharp	18												
242	t	Road	hs	f	AMR _NII	yes	Sharp	18	1,79	2,71	1,96	0,66	0,69	0,62	2,79	2,88	2,33	1,10	1,08	0,92
243	t	Road	hs	f	AMR _NIII	yes	Sharp	18	1,54	2,58	1,54	0,66	0,88	0,59	1,96	2,50	1,75	0,81	0,72	0,79
244	v	Road	hs	m	AMR _NI	no NSA	no NSA	no NSA												
245	t	Road	hs	m	AMR _NII	no NSA	no NSA	no NSA	3,17	1,63	2,13	0,64	0,65	0,54	3,25	2,00	2,33	1,03	0,66	0,64
246	v	Road	hs	m	AMR _NIII	no NSA	no NSA	no NSA												
247	v	Road	hs	m	AMR _NI	no	Smooth	9												
248	t	Road	hs	m	AMR _NII	no	Smooth	9	3,79	2,25	2,92	0,59	0,68	0,58	3,13	1,57	1,88	1,30	0,82	0,95
249	v	Road	hs	m	AMR _NIII	no	Smooth	9												
250	t	Road	hs	m	AMR _NI	no	Smooth	18	3,58	2,79	3,08	0,58	0,66	0,50	2,79	2,46	2,42	0,78	0,78	0,58
251	t	Road	hs	m	AMR _NII	no	Smooth	18	3,58	2,88	3,21	0,72	0,74	0,66	2,72	2,74	2,38	1,01	0,85	0,82
252	t	Road	hs	m	AMR _NIII	no	Smooth	18	3,71	3,00	3,38	0,75	0,42	0,58	2,31	2,21	2,09	0,80	0,98	0,78
253	t	Road	hs	m	AMR _NI	no	Sharp	9	3,75	1,96	2,63	0,61	0,62	0,58	3,58	1,71	2,04	1,06	1,08	0,81

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
254	t	Road	hs	m	AMR _NII	no	Sharp	9	3,29	1,79	2,42	0,62	0,41	0,65	2,88	2,13	2,13	0,99	0,99	0,54
255	t	Road	hs	m	AMR _NIII	no	Sharp	9	2,33	1,88	1,96	0,87	0,45	0,55	2,25	1,67	1,63	1,15	0,92	0,65
256	v	Road	hs	m	AMR _NI	no	Sharp	18												
257	t	Road	hs	m	AMR _NII	no	Sharp	18	2,96	2,50	2,46	0,69	0,66	0,72	2,46	2,46	2,15	1,25	1,10	0,68
258	t	Road	hs	m	AMR _NIII	no	Sharp	18	3,21	2,42	2,63	0,78	0,72	0,65	1,79	1,96	1,42	0,83	1,20	0,50
259	t	Road	hs	m	AMR _NI	yes	Smooth	9	4,08	2,08	3,00	0,58	0,50	0,59	3,88	1,73	2,46	1,08	0,85	0,83
260	t	Road	hs	m	AMR _NII	yes	Smooth	9	3,58	1,96	2,58	0,65	0,62	0,78	3,31	2,04	2,21	1,04	0,75	0,72
261	t	Road	hs	m	AMR _NIII	yes	Smooth	9	2,25	1,88	1,92	0,74	0,45	0,50	2,29	1,71	1,58	1,00	0,95	0,72
262	t	Road	hs	m	AMR _NI	yes	Smooth	18	4,04	2,96	3,25	0,55	0,46	0,53	3,13	2,67	2,67	1,03	0,82	0,87
263	t	Road	hs	m	AMR _NII	yes	Smooth	18	3,17	2,75	3,00	0,92	0,61	0,66	2,71	2,75	2,21	0,95	0,74	0,59
264	t	Road	hs	m	AMR _NIII	yes	Smooth	18	2,92	2,25	2,46	0,58	0,61	0,59	2,38	2,46	2,08	1,13	0,72	0,65
265	t	Road	hs	m	AMR _NI	yes	Sharp	9	3,71	1,96	2,75	0,75	0,36	0,61	3,71	2,33	2,92	0,86	0,96	0,72
266	t	Road	hs	m	AMR _NII	yes	Sharp	9	3,04	2,00	2,46	0,75	0,59	0,59	3,25	1,88	2,17	1,03	0,95	0,76
267	t	Road	hs	m	AMR _NIII	yes	Sharp	9	2,04	1,83	1,71	0,81	0,64	0,55	2,33	1,92	1,75	1,27	0,72	0,74
268	t	Road	hs	m	AMR _NI	yes	Sharp	18	2,71	2,96	2,54	0,86	0,62	0,72	2,54	2,92	2,38	1,22	0,88	1,01
269	t	Road	hs	m	AMR _NII	yes	Sharp	18	2,75	2,92	2,63	0,94	0,58	0,65	2,13	3,08	2,21	0,90	0,88	0,83
270	t	Road	hs	m	AMR _NIII	yes	Sharp	18	2,71	2,71	2,54	0,81	0,55	0,59	2,08	2,63	2,04	0,78	0,88	0,86
271	t	Office	hs	f	G722_NI	no NSA	no NSA	no NSA	4,54	4,00	4,25	0,59	0,00	0,44	4,63	3,75	3,88	0,65	0,68	0,68
272	t	Office	hs	f	G722_NII	no NSA	no NSA	no NSA	4,58	3,96	4,38	0,58	0,36	0,49	4,00	3,67	3,63	0,88	0,70	0,71
273	t	Office	hs	f	G722_NIII	no NSA	no NSA	no NSA	3,17	3,71	3,17	1,09	0,46	1,05	3,17	3,21	2,50	1,24	0,72	0,78
274	t	Office	hs	f	G722_NI	no	Smooth	9	4,58	4,17	4,42	0,58	0,38	0,50	4,53	3,88	4,08	0,77	0,45	0,50
275	t	Office	hs	f	G722_NII	no	Smooth	9	4,29	4,13	4,33	0,91	0,54	0,56	4,38	3,88	3,88	0,65	0,54	0,61
276	v	Office	hs	f	G722_NIII	no	Smooth	9												
277	v	Office	hs	f	G722_NI	no	Smooth	18												
278	v	Office	hs	f	G722_NII	no	Smooth	18												
279	t	Office	hs	f	G722_NIII	no	Smooth	18	2,75	4,04	2,79	1,15	0,55	1,10	3,58	3,75	3,50	1,21	0,74	1,06
280	t	Office	hs	f	G722_NI	no	Sharp	9	4,58	3,71	4,21	0,58	0,46	0,66	4,79	4,13	4,54	0,41	0,45	0,51
281	t	Office	hs	f	G722_NII	no	Sharp	9	4,58	4,00	4,42	0,58	0,29	0,58	4,33	4,08	4,04	0,87	0,72	0,69
282	t	Office	hs	f	G722_NIII	no	Sharp	9	3,83	3,92	3,79	0,87	0,50	0,78	3,21	3,71	2,96	0,98	0,69	0,81
283	t	Office	hs	f	G722_NI	no	Sharp	18	4,33	4,04	4,17	0,48	0,36	0,56	4,50	3,88	4,04	0,78	0,80	0,62
284	v	Office	hs	f	G722_NII	no	Sharp	18												
285	t	Office	hs	f	G722_NIII	no	Sharp	18	2,71	3,71	2,75	1,12	0,46	1,03	3,25	3,83	2,96	0,99	0,82	0,69
286	t	Office	hs	f	G722_NI	yes	Smooth	9	4,38	4,08	4,42	0,58	0,28	0,58	4,79	4,13	4,29	0,51	0,68	0,62
287	t	Office	hs	f	G722_NII	yes	Smooth	9	4,42	4,08	4,38	0,72	0,41	0,58	4,38	4,38	4,21	0,71	0,58	0,59
288	t	Office	hs	f	G722_NIII	yes	Smooth	9	2,54	4,04	2,83	1,06	0,20	0,92	2,96	3,13	2,46	1,04	0,85	0,83
289	v	Office	hs	f	G722_NI	yes	Smooth	18												
290	v	Office	hs	f	G722_NII	yes	Smooth	18												
291	t	Office	hs	f	G722_NIII	yes	Smooth	18	2,42	4,29	2,67	1,25	0,55	0,92	3,00	3,46	2,96	1,25	0,78	1,08
292	v	Office	hs	f	G722_NI	yes	Sharp	9												

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
293	v	Office	hs	f	G722_NII	yes	Sharp	9												
294	v	Office	hs	f	G722_NIII	yes	Sharp	9												
295	t	Office	hs	f	G722_NI	yes	Sharp	18	4,38	4,04	4,17	0,71	0,46	0,56	4,67	4,01	4,25	0,64	0,59	0,53
296	t	Office	hs	f	G722_NII	yes	Sharp	18	4,29	4,17	4,21	0,75	0,56	0,78	4,46	3,88	3,67	0,72	0,68	0,92
297	v	Office	hs	f	G722_NIII	yes	Sharp	18												
298	v	Office	hf	f	G722_NI	no NSA	no NSA	no NSA												
299	t	Office	hf	f	G722_NII	no NSA	no NSA	no NSA	4,33	3,04	3,58	0,56	0,75	0,78	4,25	2,50	3,13	0,94	0,72	0,80
300	t	Office	hf	f	G722_NIII	no NSA	no NSA	no NSA	3,46	2,96	3,29	0,72	0,62	0,62	3,00	2,34	2,25	1,10	0,63	0,74
301	v	Office	hf	f	G722_NI	no	Smooth	9												
302	v	Office	hf	f	G722_NII	no	Smooth	9												
303	v	Office	hf	f	G722_NIII	no	Smooth	9												
304	t	Office	hf	f	G722_NI	no	Smooth	18	3,46	3,17	3,33	0,78	0,56	0,64	3,54	2,58	2,83	1,22	0,78	0,64
305	t	Office	hf	f	G722_NII	no	Smooth	18	2,63	3,13	2,71	0,88	0,68	0,75	3,54	3,00	3,00	1,10	0,72	0,72
306	t	Office	hf	f	G722_NIII	no	Smooth	18	2,92	2,58	2,63	0,97	0,58	0,88	2,63	2,50	2,29	1,21	0,88	0,81
307	t	Office	hf	f	G722_NI	no	Sharp	9	4,25	3,46	3,71	0,74	0,59	0,75	4,08	2,67	3,21	0,88	0,82	0,66
308	v	Office	hf	f	G722_NII	no	Sharp	9												
309	v	Office	hf	f	G722_NIII	no	Sharp	9												
310	v	Office	hf	f	G722_NI	no	Sharp	18												
311	t	Office	hf	f	G722_NII	no	Sharp	18	4,08	3,29	3,71	0,93	0,55	0,81	3,75	2,71	2,75	0,99	0,81	0,74
312	t	Office	hf	f	G722_NIII	no	Sharp	18	3,00	3,08	2,83	0,83	0,50	0,70	1,67	2,54	1,75	0,87	1,02	0,79
313	t	Office	hf	f	G722_NI	yes	Smooth	9	4,54	3,00	3,67	0,51	0,51	0,70	4,42	2,75	3,17	0,65	0,68	0,56
314	v	Office	hf	f	G722_NII	yes	Smooth	9												
315	t	Office	hf	f	G722_NIII	yes	Smooth	9	3,88	3,38	3,58	0,61	0,58	0,65	3,46	2,33	2,46	1,06	0,56	0,83
316	v	Office	hf	f	G722_NI	yes	Smooth	18												
317	t	Office	hf	f	G722_NII	yes	Smooth	18	3,96	3,75	3,75	0,62	0,44	0,61	3,38	2,46	2,83	0,77	0,66	0,82
318	t	Office	hf	f	G722_NIII	yes	Smooth	18	2,88	3,42	2,96	0,80	0,65	0,75	2,88	2,63	2,25	1,15	0,88	0,85
319	v	Office	hf	f	G722_NI	yes	Sharp	9												
320	v	Office	hf	f	G722_NII	yes	Sharp	9												
321	t	Office	hf	f	G722_NIII	yes	Sharp	9	4,50	3,46	4,00	0,59	0,59	0,42	3,04	2,58	2,46	0,91	0,65	0,66
322	t	Office	hf	f	G722_NI	yes	Sharp	18	4,17	3,42	3,83	0,70	0,58	0,56	3,47	3,08	3,00	1,10	0,72	0,78
323	t	Office	hf	f	G722_NII	yes	Sharp	18	3,08	3,63	3,13	0,83	0,49	0,85	3,50	2,38	2,46	1,02	0,77	0,93
324	t	Office	hf	f	G722_NIII	yes	Sharp	18	2,88	3,50	3,00	0,80	0,72	0,78	2,63	2,63	2,29	0,77	0,82	0,69
325	t	Office	hs	m	G722_NI	no NSA	no NSA	no NSA	4,54	4,04	4,46	0,72	0,55	0,72	4,96	3,75	4,08	0,20	0,74	0,72
326	t	Office	hs	m	G722_NII	no NSA	no NSA	no NSA	4,58	4,21	4,63	0,58	0,41	0,58	4,08	3,58	3,50	1,06	0,72	0,88
327	v	Office	hs	m	G722_NIII	no NSA	no NSA	no NSA												
328	t	Office	hs	m	G722_NI	no	Smooth	9	4,54	4,58	4,63	0,72	0,50	0,49	4,75	3,79	4,13	0,53	0,51	0,45
329	t	Office	hs	m	G722_NII	no	Smooth	9	4,13	4,38	4,54	1,08	0,49	0,59	4,46	3,67	3,96	0,78	0,48	0,62
330	t	Office	hs	m	G722_NIII	no	Smooth	9	4,71	4,67	4,88	0,46	0,48	0,34	3,42	3,71	3,00	0,88	0,55	0,72
331	v	Office	hs	m	G722_NI	no	Smooth	18												

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
332	v	Office	hs	m	G722_NII	no	Smooth	18												
333	t	Office	hs	m	G722_NIII	no	Smooth	18	2,13	4,42	2,29	1,26	0,50	1,16	3,24	3,71	2,92	1,12	0,62	0,88
334	v	Office	hs	m	G722_NI	no	Sharp	9												
335	t	Office	hs	m	G722_NII	no	Sharp	9	4,58	4,38	4,67	0,58	0,49	0,48	4,71	3,63	4,00	0,46	0,49	0,51
336	t	Office	hs	m	G722_NIII	no	Sharp	9	3,75	4,38	4,08	0,94	0,49	0,83	4,00	3,29	3,42	0,93	0,81	0,72
337	t	Office	hs	m	G722_NI	no	Sharp	18	4,67	4,21	4,63	0,64	0,41	0,49	4,58	3,88	4,04	0,72	0,34	0,69
338	t	Office	hs	m	G722_NII	no	Sharp	18	4,71	4,75	4,71	0,55	0,44	0,46	4,58	3,79	4,17	0,83	0,78	0,70
339	t	Office	hs	m	G722_NIII	no	Sharp	18	4,13	4,08	4,17	0,80	0,41	0,64	2,86	3,54	3,00	0,95	0,51	0,59
340	t	Office	hs	m	G722_NI	yes	Smooth	9	4,75	4,13	4,67	0,44	0,45	0,48	4,71	4,17	4,25	0,86	0,64	0,85
341	v	Office	hs	m	G722_NII	yes	Smooth	9												
342	t	Office	hs	m	G722_NIII	yes	Smooth	9	4,00	4,29	4,21	0,88	0,46	0,51	3,75	3,38	3,00	1,19	0,77	0,88
343	t	Office	hs	m	G722_NI	yes	Smooth	18	4,25	4,46	4,25	0,68	0,72	0,94	4,75	3,88	4,21	0,53	0,80	0,59
344	t	Office	hs	m	G722_NII	yes	Smooth	18	3,96	4,29	4,17	0,95	0,62	0,76	4,25	3,87	3,92	1,15	0,54	0,93
345	t	Office	hs	m	G722_NIII	yes	Smooth	18	3,50	4,42	3,33	1,10	0,50	0,87	3,38	3,39	3,13	0,97	0,87	0,61
346	t	Office	hs	m	G722_NI	yes	Sharp	9	4,83	4,21	4,63	0,48	0,51	0,58	4,96	4,00	4,21	0,20	0,42	0,66
347	t	Office	hs	m	G722_NII	yes	Sharp	9	4,50	4,46	4,63	0,72	0,72	0,58	4,58	3,79	3,92	0,83	0,59	0,65
348	t	Office	hs	m	G722_NIII	yes	Sharp	9	3,17	4,17	3,33	1,05	0,38	0,92	3,33	3,54	2,83	1,01	0,59	0,64
349	t	Office	hs	m	G722_NI	yes	Sharp	18	4,46	4,71	4,58	0,59	0,46	0,50	4,67	3,83	4,13	0,87	0,48	0,54
350	t	Office	hs	m	G722_NII	yes	Sharp	18	4,79	4,75	4,79	0,41	0,44	0,41	4,17	3,58	3,67	1,05	0,58	0,87
351	t	Office	hs	m	G722_NIII	yes	Sharp	18	4,67	4,58	4,63	0,48	0,50	0,49	3,38	3,75	3,13	0,97	0,68	0,74
352	v	Office	hf	m	G722_NI	no NSA	no NSA	no NSA												
353	t	Office	hf	m	G722_NII	no NSA	no NSA	no NSA	4,67	3,38	4,00	0,48	0,58	0,72	4,63	2,54	3,25	0,65	1,10	0,85
354	t	Office	hf	m	G722_NIII	no NSA	no NSA	no NSA	4,17	3,25	3,63	0,64	0,68	0,71	3,17	2,54	2,29	1,09	0,59	0,55
355	t	Office	hf	m	G722_NI	no	Smooth	9	4,83	3,83	4,17	0,38	0,64	0,70	4,42	3,29	3,58	1,02	0,91	0,72
356	t	Office	hf	m	G722_NII	no	Smooth	9	4,92	3,67	4,17	0,28	0,48	0,64	4,08	2,92	3,00	0,83	0,83	0,66
357	t	Office	hf	m	G722_NIII	no	Smooth	9	4,33	3,71	3,92	0,56	0,62	0,50	4,00	2,63	2,79	1,10	0,71	0,59
358	t	Office	hf	m	G722_NI	no	Smooth	18	4,71	3,71	4,29	0,46	0,55	0,62	3,96	2,71	3,17	0,95	1,00	0,87
359	t	Office	hf	m	G722_NII	no	Smooth	18	4,75	3,79	4,13	0,53	0,72	0,68	3,96	2,96	3,04	0,62	0,62	0,69
360	t	Office	hf	m	G722_NIII	no	Smooth	18	3,00	3,50	2,96	0,66	0,59	0,62	2,75	2,54	2,25	1,07	0,83	0,90
361	t	Office	hf	m	G722_NI	no	Sharp	9	4,71	3,67	4,25	0,46	0,56	0,53	4,67	3,08	3,42	0,64	0,88	0,93
362	t	Office	hf	m	G722_NII	no	Sharp	9	4,63	3,63	4,08	0,49	0,49	0,58	4,17	3,22	3,42	0,87	0,72	0,72
363	t	Office	hf	m	G722_NIII	no	Sharp	9	2,79	3,50	2,79	0,66	0,51	0,59	3,25	2,63	2,21	1,07	0,77	0,72
364	t	Office	hf	m	G722_NI	no	Sharp	18	4,67	3,92	4,38	0,48	0,50	0,49	3,54	2,67	2,88	0,83	0,92	0,90
365	v	Office	hf	m	G722_NII	no	Sharp	18												
366	v	Office	hf	m	G722_NIII	no	Sharp	18												
367	t	Office	hf	m	G722_NI	yes	Smooth	9	4,88	3,92	4,50	0,34	0,50	0,51	4,65	3,33	3,38	0,70	0,70	0,65
368	t	Office	hf	m	G722_NII	yes	Smooth	9	4,63	3,92	4,25	0,49	0,50	0,61	3,79	3,13	3,29	1,14	0,74	0,69
369	t	Office	hf	m	G722_NIII	yes	Smooth	9	4,21	3,67	3,71	0,78	0,56	0,75	4,25	2,53	2,79	0,99	0,77	0,88
370	t	Office	hf	m	G722_NI	yes	Smooth	18	4,79	3,92	4,58	0,41	0,41	0,58	4,29	3,21	3,33	0,69	0,88	0,64

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal	Spe ech	Noi se	Glo bal
371	t	Office	hf	m	G722_NII	yes	Smooth	18	4,58	4,13	4,46	0,58	0,54	0,66	4,04	3,04	3,25	1,04	0,86	0,74
372	v	Office	hf	m	G722_NIII	yes	Smooth	18												
373	v	Office	hf	m	G722_NI	yes	Sharp	9												
374	t	Office	hf	m	G722_NII	yes	Sharp	9	3,75	3,58	3,63	0,85	0,65	0,71	4,42	2,88	3,25	0,78	0,90	0,74
375	t	Office	hf	m	G722_NIII	yes	Sharp	9	2,88	3,67	3,00	0,85	0,70	0,83	3,17	2,50	2,50	0,87	0,72	0,59
376	t	Office	hf	m	G722_NI	yes	Sharp	18	4,67	4,25	4,58	0,56	0,61	0,58	3,83	2,50	2,75	1,13	0,78	0,79
377	v	Office	hf	m	G722_NII	yes	Sharp	18												
378	v	Office	hf	m	G722_NIII	yes	Sharp	18												
379	v	Cafeteria	hs	f	G722_NI	no NSA	no NSA	no NSA												
380	t	Cafeteria	hs	f	G722_NII	no NSA	no NSA	no NSA	3,08	1,75	2,29	0,88	0,61	0,81	3,96	1,75	2,33	0,86	0,94	0,76
381	t	Cafeteria	hs	f	G722_NIII	no NSA	no NSA	no NSA	2,71	1,46	2,04	0,62	0,59	0,62	2,88	1,79	1,79	0,90	0,93	0,66
382	t	Cafeteria	hs	f	G722_NI	no	Smooth	9	3,17	1,92	2,67	0,56	0,58	0,70	4,04	2,13	2,63	0,81	0,99	0,58
383	v	Cafeteria	hs	f	G722_NII	no	Smooth	9												
384	v	Cafeteria	hs	f	G722_NIII	no	Smooth	9												
385	t	Cafeteria	hs	f	G722_NI	no	Smooth	18	2,75	2,50	2,50	0,68	0,59	0,51	3,17	2,25	2,50	0,92	0,61	0,72
386	v	Cafeteria	hs	f	G722_NII	no	Smooth	18												
387	t	Cafeteria	hs	f	G722_NIII	no	Smooth	18	2,88	2,08	2,33	0,80	0,58	0,70	2,04	1,75	1,67	0,91	0,74	0,70
388	t	Cafeteria	hs	f	G722_NI	no	Sharp	9	3,29	1,42	2,13	0,95	0,58	0,61	4,38	1,83	2,58	0,65	0,76	0,83
389	t	Cafeteria	hs	f	G722_NII	no	Sharp	9	3,08	1,29	2,08	0,78	0,46	0,58	3,54	1,91	2,38	1,06	0,88	0,65
390	v	Cafeteria	hs	f	G722_NIII	no	Sharp	9												
391	t	Cafeteria	hs	f	G722_NI	no	Sharp	18	2,83	2,04	2,21	0,82	0,62	0,72	3,00	2,25	2,25	1,06	0,90	0,79
392	v	Cafeteria	hs	f	G722_NII	no	Sharp	18												
393	v	Cafeteria	hs	f	G722_NIII	no	Sharp	18												
394	t	Cafeteria	hs	f	G722_NI	yes	Smooth	9	3,46	1,67	2,42	0,83	0,56	0,58	3,58	2,13	2,54	0,97	0,61	0,66
395	t	Cafeteria	hs	f	G722_NII	yes	Smooth	9	2,83	1,88	2,25	0,70	0,54	0,61	3,75	2,33	2,79	1,03	0,70	0,72
396	v	Cafeteria	hs	f	G722_NIII	yes	Smooth	9												
397	t	Cafeteria	hs	f	G722_NI	yes	Smooth	18	2,71	2,08	2,33	0,86	0,58	0,76	3,17	2,75	2,75	1,13	0,74	0,79
398	t	Cafeteria	hs	f	G722_NII	yes	Smooth	18	2,71	2,58	2,50	0,55	0,72	0,59	3,33	2,63	2,54	1,01	0,71	0,66
399	t	Cafeteria	hs	f	G722_NIII	yes	Smooth	18	2,67	2,33	2,50	0,64	0,56	0,51	2,63	2,17	2,04	1,06	0,64	0,69
400	t	Cafeteria	hs	f	G722_NI	yes	Sharp	9	3,04	1,63	2,42	0,69	0,58	0,72	3,83	2,29	2,71	0,92	0,75	0,55
401	t	Cafeteria	hs	f	G722_NII	yes	Sharp	9	2,63	1,75	2,04	0,77	0,61	0,55	3,75	2,04	2,33	0,68	0,86	0,82
402	t	Cafeteria	hs	f	G722_NIII	yes	Sharp	9	3,04	1,67	2,21	0,91	0,48	0,83	2,71	1,90	2,25	0,81	0,78	0,68
403	t	Cafeteria	hs	f	G722_NI	yes	Sharp	18	2,08	2,54	2,17	0,83	0,93	0,64	2,79	2,21	2,21	1,02	0,93	0,72
404	t	Cafeteria	hs	f	G722_NII	yes	Sharp	18	2,38	2,42	2,29	0,65	0,72	0,62	2,38	2,46	2,04	0,82	0,78	0,62
405	t	Cafeteria	hs	f	G722_NIII	yes	Sharp	18	2,13	2,00	1,83	0,90	0,78	0,56	2,42	2,71	2,08	1,10	1,08	0,58
406	t	Cafeteria	hs	m	G722_NI	no NSA	no NSA	no NSA	3,50	1,63	2,50	0,66	0,58	0,72	4,13	1,63	2,42	0,85	0,82	0,83
407	t	Cafeteria	hs	m	G722_NII	no NSA	no NSA	no NSA	2,58	1,67	2,00	0,78	0,56	0,59	3,21	1,58	2,20	0,78	0,65	0,58
408	t	Cafeteria	hs	m	G722_NIII	no NSA	no NSA	no NSA	1,88	1,50	1,54	0,74	0,51	0,59	2,58	1,63	1,71	1,02	0,92	0,75
409	t	Cafeteria	hs	m	G722_NI	no	Smooth	9	3,46	2,00	2,67	0,66	0,72	0,48	4,04	2,04	2,67	0,86	0,75	0,64

Condition	Subset t	Noise	Recordin	Speaker	Network	DAV	Smooth /Sharp	dB	CZECH						FRENCH					
									MOS			Standard deviation			MOS			Standard deviation		
									Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global	Speech	Noise	Global
410	t	Cafeteria	hs	m	G722_NII	no	Smooth	9	2,38	1,88	2,29	0,71	0,54	0,55	3,54	1,71	2,21	1,41	0,99	0,83
411	t	Cafeteria	hs	m	G722_NIII	no	Smooth	9	2,21	1,75	1,92	0,59	0,53	0,50	3,67	1,79	2,13	0,92	0,88	0,95
412	t	Cafeteria	hs	m	G722_NI	no	Smooth	18	3,13	2,17	2,58	0,74	0,56	0,58	3,25	2,21	2,71	0,99	0,72	0,81
413	t	Cafeteria	hs	m	G722_NII	no	Smooth	18	3,00	2,08	2,42	0,66	0,50	0,58	2,83	1,75	1,96	1,01	0,68	0,81
414	t	Cafeteria	hs	m	G722_NIII	no	Smooth	18	3,29	2,21	2,63	0,75	0,72	0,65	2,58	1,67	1,63	1,14	0,96	0,65
415	v	Cafeteria	hs	m	G722_NI	no	Sharp	9												
416	t	Cafeteria	hs	m	G722_NII	no	Sharp	9	3,21	1,46	2,29	0,72	0,59	0,55	4,00	1,92	2,38	1,14	0,83	0,97
417	t	Cafeteria	hs	m	G722_NIII	no	Sharp	9	3,17	1,96	2,42	0,87	0,69	0,65	2,50	1,78	2,13	1,29	0,72	0,74
418	t	Cafeteria	hs	m	G722_NI	no	Sharp	18	2,96	1,96	2,54	0,69	0,55	0,72	2,96	1,63	1,88	1,04	0,71	0,74
419	t	Cafeteria	hs	m	G722_NII	no	Sharp	18	2,88	1,96	2,29	0,54	0,46	0,55	3,79	1,83	2,25	1,35	0,87	0,79
420	v	Cafeteria	hs	m	G722_NIII	no	Sharp	18												
421	t	Cafeteria	hs	m	G722_NI	yes	Smooth	9	3,96	1,83	2,75	0,62	0,48	0,68	3,79	2,00	2,63	0,98	0,72	0,82
422	t	Cafeteria	hs	m	G722_NII	yes	Smooth	9	3,71	1,88	2,67	0,55	0,54	0,48	3,67	1,88	2,42	1,09	0,85	0,88
423	v	Cafeteria	hs	m	G722_NIII	yes	Smooth	9												
424	t	Cafeteria	hs	m	G722_NI	yes	Smooth	18	2,83	2,67	2,58	0,82	0,70	0,58	3,17	2,41	2,50	1,13	0,66	0,78
425	v	Cafeteria	hs	m	G722_NII	yes	Smooth	18												
426	t	Cafeteria	hs	m	G722_NIII	yes	Smooth	18	2,79	2,46	2,42	0,72	0,51	0,50	2,58	2,33	2,08	1,02	0,87	0,88
427	v	Cafeteria	hs	m	G722_NI	yes	Sharp	9												
428	t	Cafeteria	hs	m	G722_NII	yes	Sharp	9	2,29	1,79	1,79	0,81	0,72	0,66	3,58	2,04	2,33	0,88	0,91	0,76
429	t	Cafeteria	hs	m	G722_NIII	yes	Sharp	9	2,67	1,58	2,00	0,56	0,58	0,51	2,75	1,58	1,96	1,07	0,72	0,69
430	t	Cafeteria	hs	m	G722_NI	yes	Sharp	18	2,92	2,83	2,67	1,10	0,76	0,76	2,92	2,00	1,96	1,06	0,83	0,81
431	t	Cafeteria	hs	m	G722_NII	yes	Sharp	18	2,88	2,83	2,54	0,74	0,70	0,59	2,71	2,38	2,50	1,00	0,71	0,66
432	v	Cafeteria	hs	m	G722_NIII	yes	Sharp	18												

Table A.2 gives, for each subjective test condition, as defined in table A.1, the Listening level dB SPL (Active Speech Level as per P.56, no weighting) for Czech language and the level correction (dBov) for French language, according to annex B, clause B.2 Listening levels.

Table A.2: Listening level dB SPL for Czech language and level correction for French language

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
1	72,8	9,37	6,63	6,89	8,82	6,31	8,32
2	72,41	9,61	6,56	6,86	8,14	5,52	8,18
3	74,26	10,27	6,44	7,31	9,30	7,36	8,99
4	68,01	8,83	6,69	7,52	7,58	6,98	8,27
5	70,47	11,86	9,48	9,57	8,03	9,20	11,15
6	70,13	13,19	10,83	11,47	10,96	11,99	12,09
7	67,68	9,32	9,59	10,38	8,08	9,73	10,95
8	68,12	10,80	10,79	11,47	8,85	10,64	12,14
9	63,27	12,63	12,06	13,13	10,70	12,34	13,01
10	69,33	12,88	10,18	11,16	11,12	10,60	11,95
11	72,05	9,50	7,36	7,95	8,10	6,19	10,00
12	70,18	12,24	8,51	10,03	11,95	8,89	11,86
13	67,39	6,18	4,98	6,61	5,24	5,13	7,28
14	69,95	6,99	4,25	6,04	6,63	3,90	7,76
15	71,11	9,35	8,28	10,19	8,93	8,86	11,54
16	68,8	7,51	5,86	5,86	5,77	5,06	6,75
17	69,63	10,32	8,78	8,23	6,42	7,79	10,14
18	69,02	12,14	10,46	10,17	8,92	9,43	11,59
19	66,17	6,95	6,62	6,98	5,04	6,58	7,62
20	68,91	11,28	10,89	10,87	8,71	10,66	11,76
21	68,89	12,05	11,65	12,03	9,70	10,43	12,25
22	70,18	6,37	3,58	4,48	3,69	3,06	5,51
23	72,03	8,13	6,31	6,98	5,04	5,81	8,20
24	71,41	9,18	7,56	7,91	5,79	6,46	9,46
25	71,85	9,22	6,43	7,95	7,67	6,24	8,88
26	72,84	8,43	6,61	7,80	6,23	6,66	8,30
27	72,62	7,87	7,21	8,24	6,25	6,01	8,23
28	78,06	4,36	5,16	4,05	4,70	3,80	5,22
29	78,21	5,03	6,59	5,38	6,14	5,11	6,73
30	75,71	4,31	5,26	4,25	4,77	4,28	5,58
31	70,3	11,71	12,81	12,42	11,86	11,83	13,22
32	73,81	10,57	11,25	10,46	9,83	10,07	11,89
33	70,79	10,32	11,56	11,12	9,72	10,53	12,31
34	67,37	16,00	16,68	16,81	15,90	16,51	17,89
35	68,07	16,39	17,08	17,02	16,35	16,42	19,23
36	62,5	18,99	19,28	19,77	18,65	19,22	21,10
37	71,44	10,76	11,84	11,46	10,82	10,69	12,67
38	74,82	7,65	9,15	8,39	7,01	7,46	9,99
39	73	9,25	10,61	10,37	9,10	9,39	11,68
40	71,5	12,77	12,24	13,40	12,01	12,20	14,53
41	69,15	16,50	15,00	16,50	13,73	15,06	18,11
42	67,17	16,45	15,40	16,57	15,49	15,49	18,18
43	69,85	12,65	13,64	13,08	12,04	12,79	14,25
44	71,81	12,53	12,58	11,34	10,93	11,08	13,26
45	70,08	11,98	12,97	12,57	11,21	12,38	13,40
46	66,81	17,30	17,12	16,73	15,15	16,83	17,64
47	66,41	20,29	18,65	18,06	17,53	18,47	20,19
48	65,46	18,72	18,24	18,16	16,83	18,35	18,83
49	70,79	11,85	12,81	12,49	11,68	11,61	13,60
50	72,68	10,15	11,20	10,92	10,30	10,24	13,50
51	70,74	10,58	11,43	11,10	10,51	10,30	12,26
52	65,8	19,96	18,02	18,68	18,35	19,09	21,16

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
53	69,78	16,28	15,84	16,75	16,32	17,40	19,20
54	66,65	15,15	14,60	13,99	12,14	13,71	15,61
55	74,86	7,20	8,58	7,74	8,20	8,11	8,77
56	71,39	8,45	9,98	8,94	9,53	9,49	10,14
57	72,87	7,43	8,83	7,83	8,59	8,02	9,02
58	67,46	11,64	14,16	11,19	12,44	14,08	13,12
59	67,15	12,11	16,04	12,03	13,39	15,24	14,42
60	69,44	9,34	12,25	10,28	10,75	12,29	11,89
61	72,34	8,32	13,06	6,65	9,75	12,05	10,99
62	68,32	8,72	12,57	6,47	9,70	10,74	9,05
63	70,38	10,59	15,68	9,71	11,54	14,18	10,92
64	67,96	10,88	13,16	10,68	11,72	13,00	12,53
65	70,4	9,45	11,26	8,05	10,32	9,70	11,26
66	68,57	10,02	11,82	9,23	10,44	11,56	11,48
67	68,22	5,62	9,07	4,90	6,39	6,12	6,40
68	71,51	8,56	11,02	7,59	10,95	8,53	9,87
69	69,89	8,67	12,08	7,95	10,15	10,76	9,31
70	68,77	10,27	13,35	9,47	11,25	12,65	11,88
71	71,04	7,52	11,85	6,38	9,00	10,81	9,13
72	71,11	9,00	12,45	7,86	10,35	11,85	10,75
73	70,59	4,87	9,63	2,77	6,36	5,95	5,45
74	71,44	9,04	12,67	6,08	9,93	10,45	9,20
75	74,55	6,54	9,45	3,04	8,26	7,77	6,70
76	69,24	9,64	12,45	9,37	10,78	11,85	11,42
77	71,77	9,64	11,09	8,25	10,06	9,31	11,03
78	70,86	4,66	6,31	3,59	5,21	6,18	6,81
79	73,81	8,23	9,55	6,40	9,98	7,51	8,98
80	71,9	8,53	11,05	7,37	9,63	9,50	8,55
81	71,64	3,12	6,24	2,77	4,00	3,37	4,90
82	78,13	3,66	4,21	4,66	5,07	5,27	3,71
83	76,71	4,70	5,67	6,01	5,93	6,92	5,19
84	77,71	3,35	4,71	5,11	5,32	5,54	4,37
85	69,77	11,21	11,35	11,26	12,85	12,24	10,84
86	70,32	9,29	9,80	9,82	11,16	11,30	9,69
87	70,16	10,06	11,15	9,87	11,58	10,93	10,18
88	65,74	14,54	15,28	14,90	16,04	17,19	14,28
89	64,45	15,62	15,20	16,36	17,05	18,31	14,59
90	61,83	18,39	19,20	17,97	18,72	20,02	17,36
91	61,83	10,30	10,39	10,52	11,74	11,47	9,81
92	73,73	7,18	7,40	7,89	9,04	8,30	6,66
93	71,87	8,93	9,07	9,36	10,68	10,25	9,36
94	68,17	11,86	10,92	12,68	12,92	14,00	9,74
95	65,95	16,57	13,73	17,51	17,43	17,15	11,68
96	63,58	15,16	14,73	16,14	15,85	17,40	13,69
97	69,08	11,48	11,68	12,01	13,29	13,40	11,60
98	70,4	9,75	9,88	10,64	12,07	13,06	9,47
99	69,76	11,04	10,94	11,53	12,91	12,56	10,59
100	64,64	13,87	14,45	15,52	14,95	18,14	13,45
101	63,11	16,47	15,75	18,85	18,13	20,99	14,16
102	63,38	14,99	15,59	16,95	16,47	19,86	14,40
103	69,71	10,41	10,77	11,20	12,65	12,89	9,30
104	71,36	8,44	8,60	9,57	11,12	12,51	8,11
105	71,1	9,13	8,93	10,09	12,12	13,26	8,61
106	61,71	16,35	13,83	18,66	18,47	22,55	12,17
107	64,86	14,09	14,57	16,52	16,35	20,98	8,97
108	66,41	11,27	11,20	13,04	12,44	17,29	9,68
109	72,13	9,02	5,82	7,00	6,05	5,21	7,55
110	72,22	9,27	5,44	6,90	5,94	5,26	7,66

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
111	71,23	9,77	6,06	7,09	6,25	5,11	9,17
112	66,39	7,02	5,23	7,29	4,79	6,26	6,66
113	69	9,08	7,46	9,02	6,57	7,94	9,01
114	66,86	11,58	10,24	12,01	11,14	11,96	11,70
115	62,73	10,86	10,88	12,40	10,19	11,63	11,89
116	65,65	13,21	12,95	14,52	11,85	13,75	14,10
117	59,51	9,71	10,15	11,79	10,79	10,59	11,12
118	67,9	11,07	8,97	10,96	9,48	9,68	10,35
119	69,21	9,26	6,39	8,57	7,49	5,97	8,34
120	69,34	10,55	7,32	9,63	9,03	7,72	9,57
121	65,21	5,89	4,42	6,39	4,97	4,87	6,08
122	66,67	9,05	6,47	8,89	8,02	6,69	8,83
123	65,86	9,00	6,29	7,74	7,12	6,05	8,55
124	65,99	4,61	3,48	6,29	3,14	4,28	5,13
125	68,31	8,16	7,31	9,02	6,13	7,45	8,49
126	66,65	9,31	7,54	9,69	8,22	7,77	9,54
127	60,82	5,79	5,30	7,96	4,63	6,33	6,84
128	63,24	10,01	9,46	11,80	8,26	10,33	11,00
129	63,04	6,86	6,59	8,94	5,88	8,17	7,95
130	67,28	9,23	7,43	9,53	7,52	7,65	9,12
131	69,63	7,33	5,60	7,34	5,19	5,57	7,29
132	69,14	3,59	2,08	4,14	1,91	1,92	3,33
133	64,04	8,96	6,73	8,74	7,46	6,41	8,81
134	66,6	8,94	7,28	9,65	7,60	7,51	9,21
135	64,43	8,65	6,55	8,64	6,74	6,50	8,94
136	72,05	10,57	8,50	5,02	8,91	7,88	5,59
137	72,12	10,54	8,54	5,03	9,14	7,87	5,98
138	73,3	10,40	8,72	5,00	9,45	8,21	6,24
139	64,96	16,51	15,70	11,85	15,09	14,86	12,74
140	65,51	13,10	12,91	7,88	11,56	11,93	9,94
141	60,26	15,90	14,89	11,56	13,65	13,99	13,06
142	60,91	15,18	15,22	13,53	13,81	14,62	14,16
143	70,48	19,96	18,45	16,27	17,07	17,25	16,98
144	60,26	19,01	19,09	17,25	18,04	19,53	18,27
145	65,96	15,31	13,90	10,90	14,12	13,47	11,49
146	68,29	13,00	11,87	8,67	12,36	11,23	9,72
147	67,76	13,49	13,13	9,26	13,00	11,94	10,71
148	61,22	13,89	11,89	12,12	13,13	12,28	13,72
149	64,08	17,98	13,42	14,31	16,25	13,89	16,38
150	62,39	16,99	14,81	15,95	16,44	15,28	17,72
151	63,04	17,18	17,23	13,22	15,38	15,39	15,06
152	64,7	15,83	15,79	11,45	13,46	13,75	13,75
153	63,53	16,72	16,51	12,38	14,82	14,62	14,27
154	58,84	18,39	21,42	17,40	16,32	17,51	18,82
155	56,25	22,91	23,29	18,73	18,97	19,11	21,18
156	57,59	19,72	22,59	18,84	17,50	19,56	20,42
157	63,7	16,00	15,93	12,47	14,48	13,90	14,77
158	63,27	16,84	16,28	12,50	15,34	13,96	15,44
159	64,83	14,77	14,47	11,08	13,35	12,50	13,43
160	55,45	20,34	25,13	20,78	17,08	16,29	25,53
161	58,93	18,92	22,94	17,84	14,82	13,99	22,09
162	61,27	15,30	19,88	15,40	12,19	13,19	18,67
163	73,73	6,97	7,76	8,19	9,47	7,96	7,59
164	73,65	7,42	7,69	8,39	9,42	8,84	7,66
165	73,28	7,72	8,51	8,62	9,55	8,30	8,00
166	71,4	4,95	8,95	6,42	5,91	6,28	8,05
167	70,91	7,56	10,73	7,31	8,44	8,25	10,42
168	73,3	10,30	13,39	10,81	10,96	11,89	12,39

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
169	71,84	8,56	16,23	7,08	9,03	9,14	13,55
170	70,08	7,74	14,16	6,47	8,25	8,87	12,68
171	68,05	7,18	15,49	6,93	8,70	8,31	12,61
172	71,88	9,68	13,11	10,69	11,14	11,13	12,44
173	74,03	8,15	10,50	8,93	10,46	8,84	10,83
174	72,27	10,26	11,25	9,44	9,59	9,39	11,34
175	71,64	4,58	10,05	3,21	5,23	5,12	8,12
176	74,19	7,14	12,48	6,71	8,84	7,37	13,04
177	73,94	6,72	12,38	4,29	7,79	6,40	14,12
178	73,54	3,43	9,02	3,64	4,14	3,46	7,35
179	75,76	7,10	12,56	6,40	7,68	7,37	11,98
180	74,2	7,87	13,86	10,38	9,42	9,05	12,08
181	72,55	3,84	12,69	2,57	5,42	4,63	6,53
182	72,05	7,93	16,92	7,41	9,15	9,18	10,86
183	76,96	5,71	14,26	3,04	6,18	5,72	11,86
184	74	8,16	12,99	9,01	9,35	8,92	11,56
185	76,53	9,33	13,04	8,84	9,69	9,42	12,60
186	74,33	6,33	11,04	6,72	7,77	7,72	10,12
187	74,26	7,51	11,98	6,12	9,45	7,17	10,80
188	72,76	7,84	12,20	6,39	8,56	8,94	10,40
189	73,25	7,02	11,92	5,87	9,42	7,37	10,48
190	76,38	9,75	9,97	5,16	10,77	9,97	5,95
191	76,14	9,63	10,16	5,33	10,78	9,98	6,23
192	76,02	9,63	9,94	5,48	11,43	10,17	5,82
193	69,14	15,75	16,42	11,96	16,48	16,92	12,97
194	72,2	12,45	14,04	9,33	12,32	13,77	9,91
195	69,57	15,27	15,56	10,72	15,31	15,91	12,27
196	63,7	14,28	16,82	11,06	14,14	15,95	12,36
197	64,51	18,92	21,94	13,70	18,60	19,36	18,03
198	59,01	18,12	21,40	15,64	18,00	19,85	18,67
199	70,16	14,44	15,30	10,77	15,41	15,63	11,96
200	72,05	12,64	14,22	8,19	13,58	13,33	9,90
201	70,94	13,25	13,94	9,31	13,47	13,87	10,53
202	63,44	13,70	15,40	9,30	13,66	14,43	14,14
203	62,22	19,17	21,61	12,15	19,34	17,85	17,37
204	64,46	17,63	19,56	13,30	17,12	18,81	16,60
205	67,66	17,11	18,64	12,34	16,87	17,83	13,88
206	69,64	15,63	17,09	11,44	15,47	16,71	12,83
207	68,52	16,17	17,72	11,72	15,42	16,86	13,28
208	63,64	18,50	22,68	15,05	15,80	19,51	16,84
209	60,37	22,42	26,05	16,65	19,84	23,05	20,27
210	62,23	20,00	24,13	16,42	17,90	21,58	17,70
211	68,19	15,88	17,95	11,39	16,07	16,90	13,36
212	67,64	15,01	16,87	9,66	14,00	15,49	11,34
213	69,6	15,23	16,62	10,09	14,94	15,64	11,72
214	58,73	19,42	26,61	9,55	15,70	18,13	18,48
215	62,19	20,62	27,06	12,35	17,98	19,29	19,40
216	66,21	15,43	21,71	10,40	13,64	15,94	14,31
217	72	9,06	6,23	6,56	8,43	4,84	8,44
218	73,27	9,70	6,54	6,59	8,26	4,81	8,51
219	72,26	9,25	6,59	6,99	8,47	4,88	8,56
220	67,42	8,02	6,84	6,90	8,31	5,19	8,28
221	70,22	11,05	8,47	8,90	10,32	7,73	10,66
222	67,94	13,32	11,16	11,44	13,30	10,67	12,89
223	67,82	7,01	7,15	7,00	8,06	5,96	8,10
224	68,8	11,22	11,19	11,10	10,98	9,82	12,15
225	65,53	11,03	10,59	11,10	11,58	10,62	12,17
226	68,8	6,93	4,72	5,56	7,07	3,72	6,96

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
227	72,76	10,90	6,37	8,18	10,45	6,49	10,25
228	71,5	11,90	8,05	10,18	10,95	7,44	10,79
229	68,27	6,35	5,18	6,32	7,74	4,79	6,55
230	71,26	9,66	6,71	8,82	10,49	6,75	9,38
231	68,93	11,30	9,30	11,32	13,93	9,65	10,85
232	67,07	11,84	10,23	9,63	12,44	8,90	12,01
233	68,88	9,40	8,86	7,48	9,16	7,30	9,63
234	68,68	11,04	9,53	8,07	9,76	7,49	10,61
235	64,63	7,30	7,29	5,90	6,68	5,80	7,71
236	66,12	11,49	11,57	9,88	10,38	9,96	11,79
237	68,5	11,73	11,11	9,66	10,99	9,92	11,60
238	68,2	11,11	8,99	9,51	12,15	8,31	11,01
239	71,05	9,66	6,76	7,60	9,21	6,51	9,65
240	70,33	9,14	6,50	7,14	9,36	6,31	9,43
241	72,39	8,81	6,47	6,98	8,03	5,82	8,13
242	70,39	7,02	4,73	5,48	6,50	4,12	6,48
243	70,91	8,27	6,44	8,01	7,70	5,93	8,04
244	73,43	6,05	9,17	8,89	7,88	7,05	9,43
245	73,33	6,31	9,20	9,90	7,93	7,09	9,44
246	73,18	6,12	9,34	10,95	8,15	7,46	9,55
247	67	10,92	14,40	11,95	12,04	12,94	13,19
248	67,15	11,14	14,75	12,27	12,13	13,06	12,97
249	67,3	8,62	12,16	7,47	7,77	12,74	9,40
250	64,21	11,22	15,04	8,28	10,44	13,39	9,67
251	62,98	13,64	17,94	10,99	13,00	15,63	13,31
252	61,54	14,98	18,48	10,69	12,78	16,52	12,31
253	67,97	4,81	8,39	6,13	6,22	6,81	7,54
254	69,98	8,05	11,49	9,68	9,94	9,15	11,82
255	69,96	8,43	12,06	10,51	10,21	10,57	10,90
256	64,38	6,58	9,44	5,02	6,96	8,84	7,07
257	66,49	9,53	11,10	8,05	10,42	10,16	10,17
258	65,65	10,22	12,20	8,96	10,30	11,18	12,48
259	66,25	10,71	14,63	9,56	10,42	11,83	10,92
260	68,06	8,32	13,32	7,47	7,89	9,51	9,34
261	67,46	9,22	12,87	8,08	8,30	10,46	10,30
262	59,86	8,19	12,69	2,76	5,54	7,91	4,83
263	58,72	12,00	15,71	7,04	8,94	11,04	10,26
264	62,67	9,61	12,63	4,38	6,66	9,19	6,62
265	66,13	10,77	14,30	9,77	10,56	11,59	11,41
266	68,98	7,98	10,81	6,57	7,55	8,86	9,18
267	68,07	9,25	12,39	9,29	8,10	9,37	10,07
268	62,76	9,20	10,94	6,24	8,79	7,46	8,21
269	69,86	8,54	10,49	5,51	7,16	8,20	7,80
270	64,29	9,40	11,64	6,63	9,54	8,57	8,35
271	74,23	10,68	7,87	8,96	8,82	7,57	9,85
272	72,73	4,55	1,78	2,13	1,84	1,71	3,23
273	71,67	9,08	6,43	7,67	6,63	6,25	8,52
274	72,49	8,30	5,94	7,58	6,52	5,86	8,15
275	72,35	8,22	6,70	8,12	5,80	7,18	7,87
276	73,68	7,01	5,33	6,73	5,48	5,28	7,11
277	73,06	9,41	8,29	10,25	7,65	8,68	9,88
278	73,21	9,03	6,80	8,68	7,45	6,87	9,05
279	73,47	8,83	7,48	8,75	6,56	7,53	8,87
280	75,22	7,92	5,35	6,88	6,28	5,12	7,45
281	72,87	8,06	5,99	7,11	6,48	5,29	7,67
282	73,6	6,17	5,12	6,28	4,89	4,73	7,26
283	72,64	9,57	7,70	9,27	7,89	7,67	9,65
284	72,95	8,06	5,99	7,11	6,48	5,29	7,67

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	**Level correction (dBov)**					
Condition		**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
285	74,77	7,71	5,34	6,74	5,90	5,22	8,10
286	74,81	7,57	5,05	6,50	5,82	4,94	7,09
287	72,15	8,28	5,23	6,67	6,34	4,75	7,32
288	73,2	9,63	7,70	9,58	7,56	7,94	9,44
289	73,77	6,41	4,89	6,42	4,67	5,11	6,62
290	73,54	7,47	6,10	6,14	4,71	6,17	7,76
291	74,05	7,75	5,81	6,81	5,05	5,55	7,35
292	75,57	7,40	4,86	6,23	5,79	4,59	6,84
293	75,19	8,89	5,58	6,95	6,56	5,38	7,51
294	75,22	6,17	4,24	4,49	3,97	3,31	5,64
295	75,24	7,47	4,93	6,51	5,83	4,70	6,91
296	74,69	7,51	5,58	6,93	5,05	5,62	7,08
297	72,38	7,06	5,81	6,28	5,05	4,53	6,94
298	63,78	14,85	11,56	13,60	14,29	13,06	14,90
299	67,86	13,15	9,63	11,55	12,54	11,15	11,95
300	68,56	12,25	8,51	10,07	11,01	11,14	10,75
301	66,6	13,52	10,62	12,96	12,41	12,62	12,61
302	68,02	10,22	7,29	9,09	8,10	9,03	10,03
303	66,77	14,16	10,83	12,97	13,95	11,65	11,64
304	63,83	14,92	13,75	15,66	13,71	15,58	15,17
305	65,62	15,36	12,65	15,03	12,21	14,50	14,24
306	64,03	15,22	12,82	15,42	13,60	15,31	14,54
307	68,91	12,09	8,45	10,71	9,60	10,22	10,59
308	70,11	12,42	10,53	11,96	10,51	12,03	11,45
309	66,04	12,34	9,32	11,56	11,02	11,50	12,59
310	65,51	11,73	8,46	11,40	9,76	10,59	10,93
311	66,85	12,32	10,32	11,93	10,78	11,96	11,45
312	66,33	12,76	10,29	12,79	9,37	13,87	13,43
313	68,01	11,19	8,99	10,93	9,04	10,79	10,92
314	65,6	13,63	10,70	13,01	11,44	13,10	13,32
315	65,7	12,31	10,21	11,79	10,57	11,25	11,69
316	67,24	11,80	9,72	11,69	9,18	11,97	10,82
317	66,2	12,88	10,85	12,46	10,22	12,65	12,91
318	64,63	13,05	10,79	12,31	8,27	12,75	12,94
319	68,32	6,15	4,52	5,77	4,60	4,52	6,14
320	67,72	11,80	9,34	11,44	11,91	10,82	11,20
321	65,73	10,57	9,37	10,58	9,23	9,99	11,89
322	69,25	12,43	9,25	11,39	10,40	10,80	11,73
323	67,38	9,55	8,17	10,14	8,32	9,47	11,16
324	68,7	11,22	9,20	11,28	8,38	10,75	10,27
325	75,74	9,00	11,58	8,24	12,07	8,53	10,43
326	74,67	6,19	8,79	5,01	8,89	5,52	7,35
327	73,86	6,04	9,28	5,42	10,41	5,62	7,45
328	74,1	6,46	7,70	5,48	8,44	6,09	7,69
329	73,98	6,52	8,34	5,23	8,04	7,16	7,42
330	71,33	5,53	7,49	5,84	7,36	5,23	6,66
331	72	6,88	8,00	5,88	8,46	6,53	8,04
332	73,66	7,62	7,96	6,95	9,39	6,71	8,86
333	75,12	7,33	7,82	5,95	8,86	6,42	7,46
334	75,41	5,11	7,01	4,15	7,62	4,66	6,21
335	72,44	6,14	8,49	5,58	8,99	5,94	7,74
336	74,73	6,69	9,01	5,01	9,15	5,73	7,40
337	71,98	6,70	7,92	5,77	8,88	6,28	7,87
338	73,78	6,14	8,48	5,58	9,03	5,94	7,74
339	73,17	6,74	9,25	5,97	8,08	7,39	7,50
340	75,37	5,97	8,20	5,04	8,16	5,52	7,11
341	74,25	6,23	9,72	4,62	8,00	6,97	7,07
342	74,51	5,08	7,15	4,33	7,62	5,59	5,61

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
343	74,52	6,01	7,68	5,09	8,12	5,54	7,12
344	74,47	6,21	8,23	5,01	7,77	7,09	6,85
345	71,86	6,52	8,19	5,30	7,80	4,40	8,81
346	75,38	5,95	8,28	5,01	8,86	5,51	7,09
347	74,36	6,26	7,97	5,33	8,91	5,46	7,45
348	74,36	5,12	9,14	3,84	7,89	5,85	7,10
349	74,55	5,97	8,20	5,02	8,14	5,52	7,11
350	74,32	6,11	8,41	5,19	8,79	5,59	7,29
351	75,26	6,34	9,52	4,74	8,10	6,60	7,30
352	67,97	14,13	17,03	15,77	16,74	15,32	17,16
353	68,18	12,54	15,38	13,49	14,16	14,85	13,62
354	69,13	10,45	14,43	13,15	14,33	11,22	13,45
355	67,44	11,90	15,08	13,60	14,55	12,23	12,58
356	70,69	12,11	14,38	13,90	14,86	12,41	13,47
357	69,9	10,94	14,02	12,26	13,04	12,30	11,64
358	67,37	13,16	16,71	14,13	15,48	15,00	12,67
359	68,51	12,93	16,04	13,04	15,68	13,71	13,95
360	67,97	14,58	16,70	13,78	14,88	13,87	12,98
361	70,54	10,34	13,62	12,07	13,25	10,95	11,38
362	71,45	11,73	14,33	12,95	14,48	13,30	13,44
363	69,64	11,89	14,93	13,09	12,81	12,34	13,83
364	66,93	10,47	13,31	11,91	13,26	10,84	11,42
365	68,57	11,80	14,47	13,58	14,58	13,09	12,53
366	69,68	11,93	14,86	12,97	13,99	13,19	10,83
367	69,88	10,99	14,46	12,33	13,99	11,57	12,51
368	69,48	12,00	15,21	13,43	14,43	13,00	12,84
369	69,3	12,54	16,70	13,57	14,42	14,53	13,59
370	70,24	11,16	14,74	11,24	14,18	11,95	12,06
371	69,33	11,47	14,98	11,42	14,47	11,82	11,97
372	69,01	10,94	14,43	11,06	13,34	12,23	11,56
373	70,68	10,92	14,13	12,27	13,72	11,49	12,13
374	69,64	11,27	14,33	12,27	14,07	11,89	12,30
375	70,53	10,86	13,70	11,12	13,27	12,16	13,54
376	69,67	9,73	13,40	9,71	11,96	11,53	10,42
377	68,6	11,25	14,69	11,16	14,29	11,92	12,68
378	70,34	10,43	13,75	10,93	14,09	12,28	12,34
379	69,94	9,67	7,56	7,56	9,83	7,17	9,48
380	69,15	13,59	10,98	11,47	13,81	11,15	13,42
381	73,1	8,21	6,19	6,58	7,80	5,46	7,78
382	68,86	11,62	9,03	9,34	10,88	9,22	10,75
383	74,1	11,68	9,07	9,48	10,80	8,67	10,98
384	70,95	12,21	10,22	10,98	12,44	10,29	11,90
385	70,71	14,31	10,52	11,86	10,83	11,86	11,91
386	70,02	14,57	10,44	11,53	10,80	11,90	11,62
387	69,22	15,67	11,61	12,03	12,21	13,02	12,73
388	74,31	8,59	7,18	7,30	8,85	7,14	8,73
389	73,22	10,84	7,55	8,49	10,33	7,78	9,67
390	72,13	11,25	7,42	8,22	9,45	7,67	10,04
391	70,61	12,12	7,75	9,66	10,68	9,06	9,70
392	71,44	10,84	7,73	8,44	10,32	7,89	9,55
393	72,13	12,29	6,44	8,46	10,37	9,35	9,25
394	72,84	12,20	9,24	9,20	9,54	8,81	9,47
395	72,69	12,71	9,42	9,47	9,74	8,99	9,67
396	69,49	13,38	12,15	12,09	12,84	12,14	12,76
397	73,07	12,78	11,37	10,85	8,83	10,12	10,00
398	71,89	13,42	11,49	11,00	8,97	10,42	10,36
399	73,15	13,80	13,69	12,48	9,97	12,12	11,46
400	73,24	10,64	6,98	8,05	9,06	7,37	8,23

	CZECH	FRENCH					
	Listening level dB SPL (Active Speech Level as per P.56, no weighting)	Level correction (dBov)					
Condition		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
401	71,98	10,23	7,43	8,41	8,40	8,31	9,04
402	69,81	9,51	7,32	8,11	7,88	7,59	8,16
403	75,43	9,65	7,51	8,18	6,60	7,31	8,05
404	73,42	11,01	7,77	8,80	7,75	5,99	8,45
405	75,18	11,17	7,72	8,45	6,69	7,36	8,18
406	70,97	8,03	8,59	8,88	8,32	8,87	8,73
407	72,13	12,07	12,56	13,37	12,38	12,93	13,06
408	70,62	7,09	7,28	7,51	7,61	7,85	7,30
409	69,39	8,83	10,68	10,01	9,99	10,52	10,96
410	70,23	5,50	7,08	6,84	5,85	7,38	6,76
411	68,03	8,68	10,19	10,26	9,47	11,05	11,15
412	67,01	9,37	13,79	9,49	12,55	12,69	13,90
413	67,13	10,37	13,87	10,89	13,22	12,69	14,37
414	66,79	10,00	13,71	10,58	13,23	11,14	12,69
415	72	7,21	8,15	7,79	7,29	8,94	7,78
416	71,05	8,79	10,29	9,60	9,24	9,74	10,37
417	69,62	8,86	9,98	9,64	9,28	9,98	9,63
418	66,26	9,40	12,16	10,24	11,43	11,24	12,75
419	67,24	8,80	10,11	9,67	9,20	9,87	10,41
420	68,88	9,22	11,60	11,03	9,83	11,66	10,55
421	70,45	8,53	13,12	9,56	11,80	10,65	12,60
422	69,45	10,31	14,50	11,04	13,30	11,51	13,80
423	69,36	9,07	12,64	10,65	11,64	10,93	11,97
424	69,35	9,25	19,29	8,06	15,47	9,19	16,50
425	72,84	9,45	19,73	8,08	15,41	9,23	15,74
426	69,89	9,72	15,44	8,15	15,05	10,72	15,63
427	70,89	8,26	11,95	9,16	11,12	9,71	12,01
428	68,74	8,23	11,85	9,40	11,58	10,11	12,27
429	69,53	8,22	11,77	9,18	9,47	10,36	11,28
430	69,94	7,84	14,40	6,71	12,40	8,85	12,93
431	69,56	9,66	16,44	9,49	15,53	9,08	15,65
432	69,19	8,12	15,67	7,12	11,96	8,26	14,27

Annex B (informative): Complementary information on the practical subjective test procedures

Clause 8.2 gives the overall procedures implemented for the subjective test. This clause describes in more details the equipment and the different steps of the process, as used by the two Laboratories involved in the subjective tests. This additional information is needed by the Laboratories in charge of the development and the validation of objective models. It also could be useful for Laboratories aiming to use the data base for future developments.

B.1 "Headphones"

Both laboratories have used binaural "Hi-Fi" Headphones.

The type and the sensitivity of these headphones are indicated below:

- **France Telecom:** Sony MDR CD1000, 104 dB/mW.
- **Mesaqin:** Sennheiser HD600 (without additional filtering), 97 dB/mW.

NOTE: FT provided the frequency responses for the eight headphones in an excel file "headphones.xls", which is available in STF 294 archives.

B.2 Listening Levels

During the discussions between the experts of the STF 294, it appeared necessary to summarize in the present document, the process implemented by each Laboratory to assess and to calibrate the listening levels of the Speech/noise sequences. The information relative to each subjective test condition is available in table A.2.

B.2.1 Mesaqin Laboratory

Experts of Mesaqin Laboratory have operated as follows.

The playback chain (digital player-distribution amplifier-headphones) has been calibrated in a way that for -26 dBov active speech level (P.56 WideBand (WB)) the listening level equals to 79 dB SPL, on the 4 original samples. This setting has been kept during the entire tests.

Consequently, the listening levels varied depending on the level of processed database files. The maximum signal level was 78,2 dB SPL, the average signal level was 69,5 dB SPL, the minimum signal level was 55,5 dB SPL. The calibration was done with artificial ear, according to ITU-T Recommendation P.57 [30], Type 3-2 at a 5 N force.

From each recording processed by the Experts from Telefonica (see Test Plan directory), **one** single sentence has been used (always different one, of course, not to make the listeners tired) and played this sentence three times to our listeners etc. as described, according to clause 8.2.3. Mesaqin experts did not introduce change in level of the files.

B.2.2 France Telecom Laboratory

Experts from France Telecom equalized the levels of all the final samples to -26 dBov, and equalized the listening levels to 79 dB SPL.

The calibration was done on B&K HATS (ITU-T Recommendation P.58 [31], implementing Type 3.3 artificial ears, as defined in ITU-T Recommendation P.57, No filter was applied (except filters which cut frequencies below 22,4 Hz and above 22,4 kHz on the pre-amplifier 2 636 B&K).

The amplification factors are available in a table in the STF 294 Archives.

Calibrating signals so that their overall level is, for all of them, 79 dB SPL (with the limit that the signal calibration should rely on clean speech). A listening check of the whole corpus was subsequently performed in order to check that the calibration on noisy speech worked well and that no big overall differences were audible.

The different processings applied on the original files introduced up to 18 dB of level differences.

From a file *level_variability.xls* (not included in the present document, but available in STF 294 archives), it should be possible to check the level variability for the sample n°1 of the female language. One figure showing the mean RMS levels (before equalization), averaged on the three network conditions and on the five noises, and the other figure showing the mean RMS levels (before equalization) according to the NSA and recording conditions for the three noises for which the two recording conditions were presented, it can be seen that the initial difference between the two recording conditions disappear, according to the noise, the NSA, etc. And the differences were rather extensive.

It can be assumed, the level of the complete sequence including the noise, measured with P.56 is always 79 dB SPL.

Note that the validation database is taken from the overall subjective database, which contains all the scores obtained with the listening tests.

B.2.3 Additional remarks

The approach chosen by the STF 294 to define the subjective data base is rather new. It appears that some clarifications need to be included to improve the definition of the process defined in ITU-T Recommendation P.835 [25], in particular when test sequences include high acoustic noise level. The calibration process needs to be clarified in ITU-T Recommendation P.835 [25].

This could take benefit from the work done by STF 294 experts, when discussing the detailed methodology.

B.3 Sample Shortening / Extraction

The Experts have chosen different approaches in the global process of the sentence used in the subjective tests

France Telecom Experts used for each network condition 6 different sentences, each being evaluated by 4 listeners, and providing average score per all 6 sentences.

Mesaqin Experts used one sentence per network condition, and this was evaluated by 24 listeners.

Annex C (informative): Instructions and Questions presented to listeners during subjective tests

Instructions to Czech listeners The following instructions have been given to subjects before the listening tests: "Nyní uslyšíte serii nahrávek. Každou z nich uslyšíte vždy třikrát po sobě, oddělené krátkou pauzou. Při prvním přehrání se soustřeďte pouze na řeč ve vzorku (respektive pouze na šum v pozadí nahrávky), při druhém pouze na šum v pozadí nahrávky (respective pouze na řeč) a ve třetím na celkovou kvalitu vzorku. To, na co jste se zaměřili, ohodnoťte ihned po příslušném přehrání celým číslem 1 až 5 dle stupnic, které vidíte před sebou".

Table C.1: Czech Questions version 1 (order Speech-Noise-Overall)

A. Zaměřte se pouze na řeč a vyberte jedno z následujících hodnocení:	**B.** Zaměřte se pouze na šum v pozadí nahrávky a vyberte jedno z následujících hodnocení:	**C.** Vyberte kategorii, nejlépe odpovídající celkové kvalitě vzorku z hlediska každodenní hlasové komunikace.
Řeč v tomto vzorku byla:	**Pozadí v tomto vzorku bylo:**	**Celková kvalita tohoto vzorku byla**
5 – nezkreslená 4 – jen mírně zkreslená 3 – poněkud zkreslená 2 – zkreslená 1 – velmi zkreslená	5 – neslyšitelné 4 – slyšitelné, ale nerušící 3 – mírně rušící 2 – rušící 1 – velmi rušící	5 – vynikající 4 – dobrá 3 – uspokojivá 2 - špatná 1 – velmi špatná

Table C.2: Czech Questions version 2 (order Noise-Speech-Overall)

A. Zaměřte se pouze na šum v pozadí nahrávky a vyberte jedno z následujících hodnocení:	**B**. Zaměřte se pouze na řeč a vyberte jedno z následujících hodnocení:	**C.** Vyberte kategorii, nejlépe odpovídající celkové kvalitě vzorku z hlediska každodenní hlasové komunikace.
Pozadí v tomto vzorku bylo:	**Řeč v tomto vzorku byla:**	**Celková kvalita tohoto vzorku byla**
5 – neslyšitelné 4 – slyšitelné, ale nerušící 3 – mírně rušící 2 – rušící 1 – velmi rušící	5 – nezkreslená 4 – jen mírně zkreslená 3 – poněkud zkreslená 2 – zkreslená 1 – velmi zkreslená	5 – vynikající 4 – dobrá 3 – uspokojivá 2 - špatná 1 – velmi špatná

Instructions to French listeners: The following instructions have been given to subjects before the listening test: "Vous allez entendre à travers le casque qui est placé devant vous des séquences d'échantillons de parole. Chaque séquence est constituée de trois phrases identiques (d'une durée de 4 secondes), séparées par un temps de réponse de 4 secondes. Pour chaque séquence, vous allez évaluer:

1) après la 1ère écoute de la phrase: dans quelle mesure la parole est déformée;

2) après la 2ième écoute de la phrase: dans quelle mesure le bruit de fond est gênant;

3) après la 3ième écoute de la phrase: la qualité globale.

Pendant l'écoute des phrases, le bouton rouge qui est devant vous sera allumé. Dans le même temps, il vous est indiqué sur l'afficheur ce que vous devez juger à la fin de cette phrase: DEFORMATION DU SIGNAL DE PAROLE à la 1ère écoute, NIVEAU DE GENE DU BRUIT DE FOND à la 2ième écoute, QUALITE GLOBALE DU SIGNAL à la 3ième écoute. Vous voudrez bien écouter chaque phrase complètement. Puis, quand le bouton vert s'allumera, les cinq boutons numérotés de 1 à 5, sur lesquels vous pouvez donner votre opinion, clignoteront.

Lors de la **première écoute**, vous donnerez votre opinion sur **la déformation du signal de parole** en appuyant sur le bouton approprié (chiffres 1, 2, 3, 4, 5) selon l'échelle suivante:

5: PAS DÉFORMÉ

4: PEU DÉFORMÉ

3: MOYENNEMENT DÉFORMÉ

2: DÉFORMÉ

1: TRES DÉFORMÉ

Lors de la **deuxième écoute,** vous donnerez votre opinion sur **le niveau de gêne du bruit de fond** en appuyant sur le bouton approprié (chiffres 1, 2, 3, 4, 5) selon l'échelle suivante:

5: INAUDIBLE

4: AUDIBLE MAIS PAS GÊNANT

3: UN PEU GÊNANT

2: GÊNANT

1: TRES GÊNANT

Lors de la **troisième écoute,** vous donnerez votre opinion sur **la qualité globale** en appuyant sur le bouton approprié (chiffres 1, 2, 3, 4, 5) selon l'échelle suivante:

5: EXCELLENTE

4: BONNE

3: MOYENNE

2: MEDIOCRE

1: MAUVAISE

Vous disposez de 5 secondes pour enregistrer votre réponse. Ce test qui comprend 4 séances, commencera par un apprentissage. Merci pour votre participation et bon courage."

NOTE: The order of the two first scales was reversed for the other half of subjects, according to ITU-T Recommendation P.835 [25].

Annex D (informative): Bibliography

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

http://www.speex.org

http://www.gnu.org

http://www.fsf.org

http://snad.ncsl.nist.gov/itg/nistnet

http://www.ethereal.com

History

Document history		
V1.1.1	July 2006	Membership Approval Procedure MV 20060915: 2006-07-18 to 2006-09-15
V1.1.1	September 2006	Publication