

3GPP TSG-S4 Meeting #27  
Munich, Germany, 7-11 July 2003

*Tdoc S4-030564*

---

**Source:** TSG SA WG4  
**Title:** Test Plan for the AMR Narrow-Band Packet switched Conversation test  
**Document for:** Information  
**Agenda Item:** 7.4.1

---

## 1. Introduction

This document contains the test plan of one conversation test for the Adaptive Multi-Rate Narrow-Band (AMR-NB) in Packet Switched networks.

All the laboratories participating to this conversation test phase will use the same test plan, just the language of the conversation would change. Even if the test rooms or the test equipments are not exactly the same in all the laboratories, the calibration procedures and the tests equipment characteristics and performance (as defined in this document) will guarantee the similarity of the test conditions.

Section 2 gives references, conventions and contacts, section 3 details the test methodology, including test arrangement and test procedure, and section 4 defines the financial considerations.

Annex A contains the instructions for the subjects participating to the conversation tests.

Annex B contains the description of results to be provided to the Analysis Laboratory (if any) by the testing laboratories.

Annex C contains the list of statistical comparisons to be performed.

Considerations about IPV6 versus IPV4 are given in section 3.2. RoHC is not implemented in AMR-NB conversation test. The effect of RoHC should be extrapolated from the results observed in AMR-WB conversation test.

## 2. References, Conventions, and Contacts

### 2.1 Permanent Documents

ITU-T Rec.P.800	Methods for Subjective Determination of Transmission Quality	
ITU-T Rec. P.831	Subjective performance evaluation of network echo cancellers	This Recommendation defines conversation test procedures based on handset telephones, and gives inputs for the calibration.

### 2.2 Key Acronyms

AMR-NB	Adaptive Multi-Rate Narrowband Speech Codec
AMR-WB	Adaptive Multi-Rate Wide-band Speech Codec
MOS	Mean Opinion Score

## 2.3 Contact Names

The following persons should be contacted for questions related to the test plan.

Section	Contact Person/Email	Organisation	Address	Telephone/Fax
Experiments and results analysis	J-Y Monfort	France Telecom R&D	2, Avenue P. Marzin, 22307 Lannion Cédex France	Tel : +33296053171 Fax : +33296051316
AOB	Paolo paolo.usai@etsi.fr	Usai ETSI MCC	650 Route des Lucioles 06921 Sophia Antipolis France	Tel: 33 (0)4 92 94 42 36 Fax: 33 (0)4 93 65 28 17

## 2.4 Responsibilities

Each test laboratory has the responsibility to organize its conversation tests.

The list of Test laboratories participating to the conversation test phase.

Lab	Company	Language	Statistical analysis	Reporting
1	LAB1			
2	LAB2			

## 3. Test methodology

### 3.1 Introduction

The protocol described below evaluates the effect of degradation such as delay and dropped packets on the quality of the communications. It corresponds to the conversation-opinion tests recommended by the ITU-T P.800 [1]. First of all, conversation-opinion tests allow subjects passing the test to be in a more realistic situation, close to the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are suited to assess the effects of impairments that can cause difficulty while conversing (such as delay).

Subjects participate to the test by couple; they are seated in separate sound-proof rooms and are asked to hold a conversation through the transmission chain performed by means of UMTS simulators and communications are impaired by means of an IP impairments simulator part of the CN simulator and by the air interface simulator, as the figure below describes it.

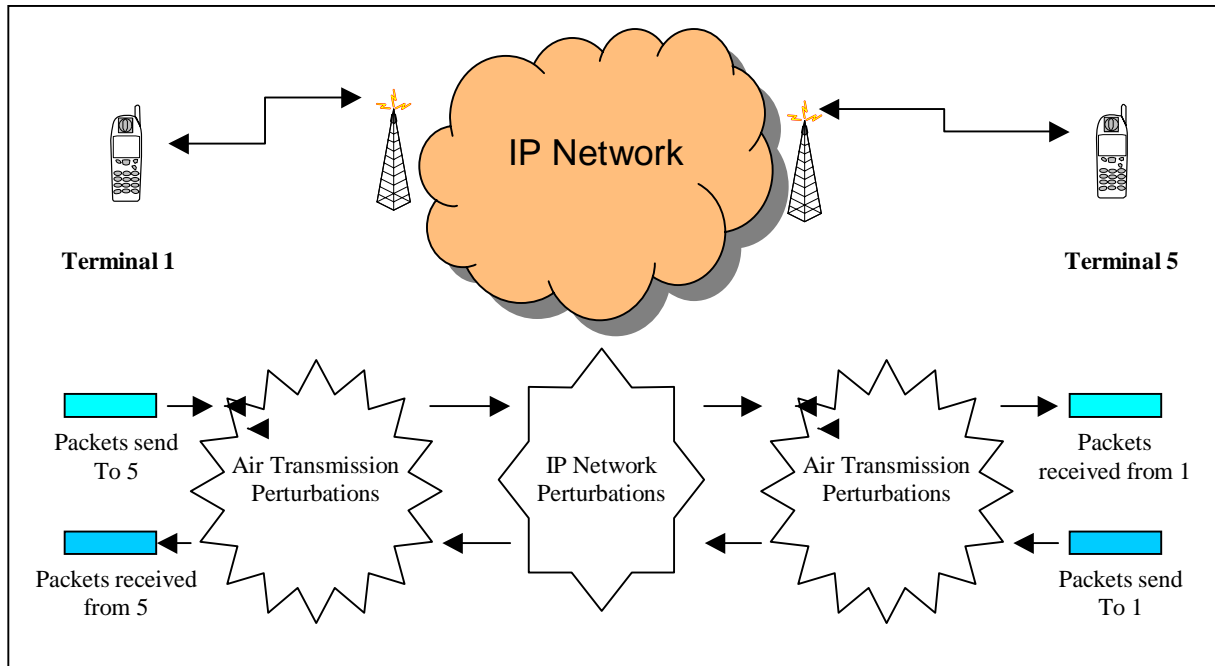
The network configurations (including the terminal equipments) will be symmetrical (in the two transmission paths). The only dissymmetry will be due to presence of background noise in one of the test rooms.

### 3.2 Test arrangement

#### 3.2.1 Description of the proposed testing system

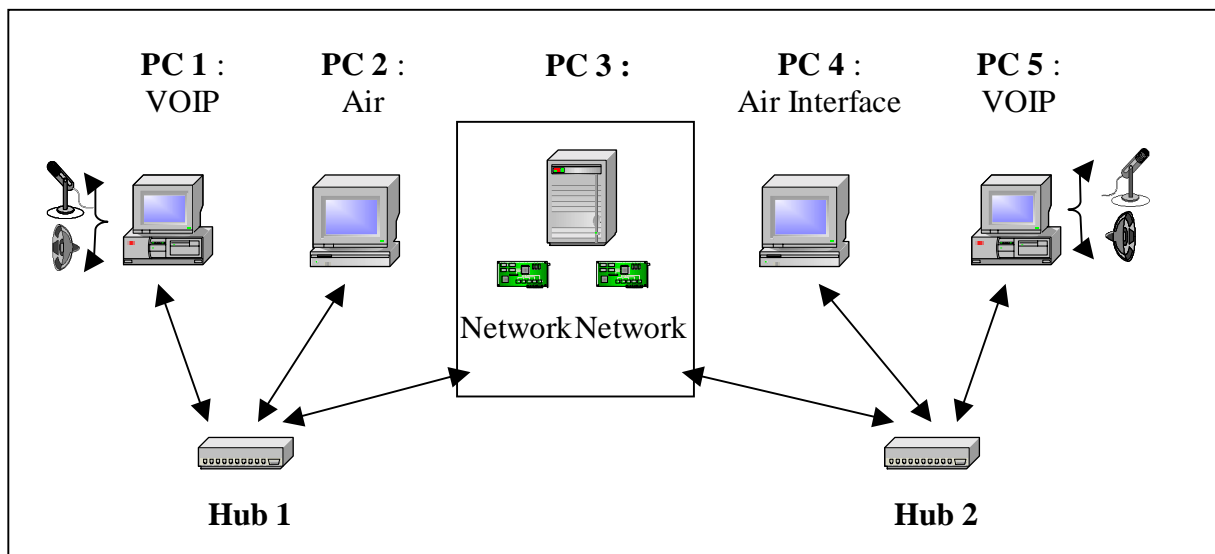
This contribution describes a UMTS simulator for the characterization of the AMR speech codecs when the bitstream is transmitted over a PS network. The procedure to do the conversational listening test has been earlier described in [1].

Figure 1 describes the system that is going to be simulated:



**Figure 1:** Packet switch audio communication simulator

This will be simulated using 5 PCs as shown in Figure 2.



**Figure 2:** Simulation Platform

PC 1 and PC 5: PCs under Windows OS with VOIP Terminal Simulator Software of France Telecom R&D.

PC 2 and PC 4: PCs under Linux OS with Air Interface Simulator of Siemens AG.

PC 3: PCs under WinNT OS with Network Simulator Software (NetDisturb).

Basic Principles:

The platform simulates a packet switch interactive communication between two users using PC1 and PC5 as their relatives VOIP terminals. PC1 sends AMR encoded packets that are encapsulated using IP/UDP/RTP headers to PC5. PC1 receives these IP/UDP/RTP audio packets from PC5.

In fact, the packets created in PC1 are sent to PC2. PC2 simulates the air interface Up Link transmission and then forwards the transmitted packets to PC4.

In the same way, PC4 simulates the air interface Down Link transmission and then forwards the packets to PC5. PC5 decodes and plays the speech back to the listener.

### 3.2.2 France Telecom Network simulator

The core network simulator, as implemented, works under IPv4.

However, as the core network simulator acts only on packets (loss, delay,...) the use of Ipv4 or Ipv6 is equivalent for this test conversation context. Considering the networks perturbations introduced by the simulator and the context of the interactive communications, the simulation using IPv4 perturbation network simulator is adapted to manage and simulate the behaviours of an IPv6 core network.

Figure 3 shows the possible parameters that can be modified.

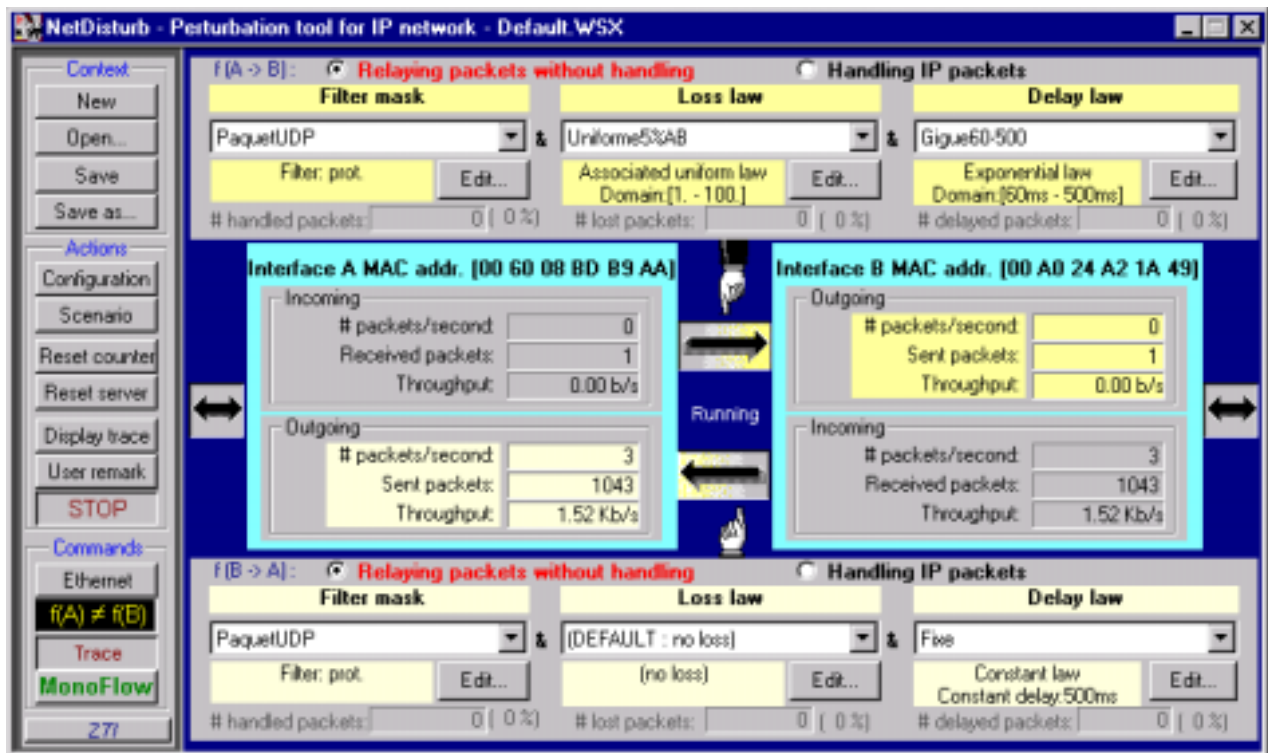


Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or on the same way. For example, delay can be set to a fixed value but can also be set to another law such as exponential law.

### 3.2.3 UMTS simulator choices

The transmission of IP/UDP/RTP/AMR packets over the UMTS air interface is simulated using the RAB described in Section 3.2.3.1 The required functions of the RLC layer are implemented according to TS 25.322 and work in real-time. The underlying Physical Layer is simulated offline. Error patterns of block errors (i.e. discarded RLC PDUs) are inserted in the real-time simulation

as described in Section 3.2.3.2 For more details on the parameter settings of the Physical Layer simulations see Section 3.2.3.3

### 3.2.3.1 RAB and protocols

For our conversational tests, the AMR will encode speech at a maximum of 12.2 kbit/s. The bitstream will be encapsulated using IP/UDP/RTP protocols. The air interface simulator will receive IPv4 (or IPv6) packets from the CN simulator. The RTP packets will be extracted and before transmission over the air interface, IPv6 headers will be inserted. Finally real IPv6 packets are transmitted over the air interface simulator.

The payload Format should be the following:

- RTP Payload Format for AMR-NB (RFC 3267) will be used;
- Bandwidth efficient mode will be used;
- One speech frame shall be encapsulated in each RTP packet;
- Interleaving will not be used;

The payload header will then consist of the 4 bits of the CMR (Codec Mode Request). Then 6 bits is added for the ToC (Table of Content). For IPv4, this corresponds to a maximum of 72 bytes per frame that is to say 28.8 kbit/s, this goes up to 92 bytes (36.8 kbit/s) when using IPv6 protocol on the air interface.

RTCP packets will be sent. However, in the test conditions defined in the conversation test plans, RTCP is not mandatory, as it is not in a multicast environment (see IETF rfc 1889) we are not going to make use of the RTCP reports.

ROHC is an optional functionality in UMTS. In order to reduce the size of the tests and the number of condition ROHC algorithm will not be used for AMR-NB conversation test. This functionality will only be tested in the wideband condition. The Conversational / Speech / UL:42.8 DL:42.8 kbps / PS RAB RAB coming from TS 34.108 v4.7.0 will be used:

Here is the RAB description:

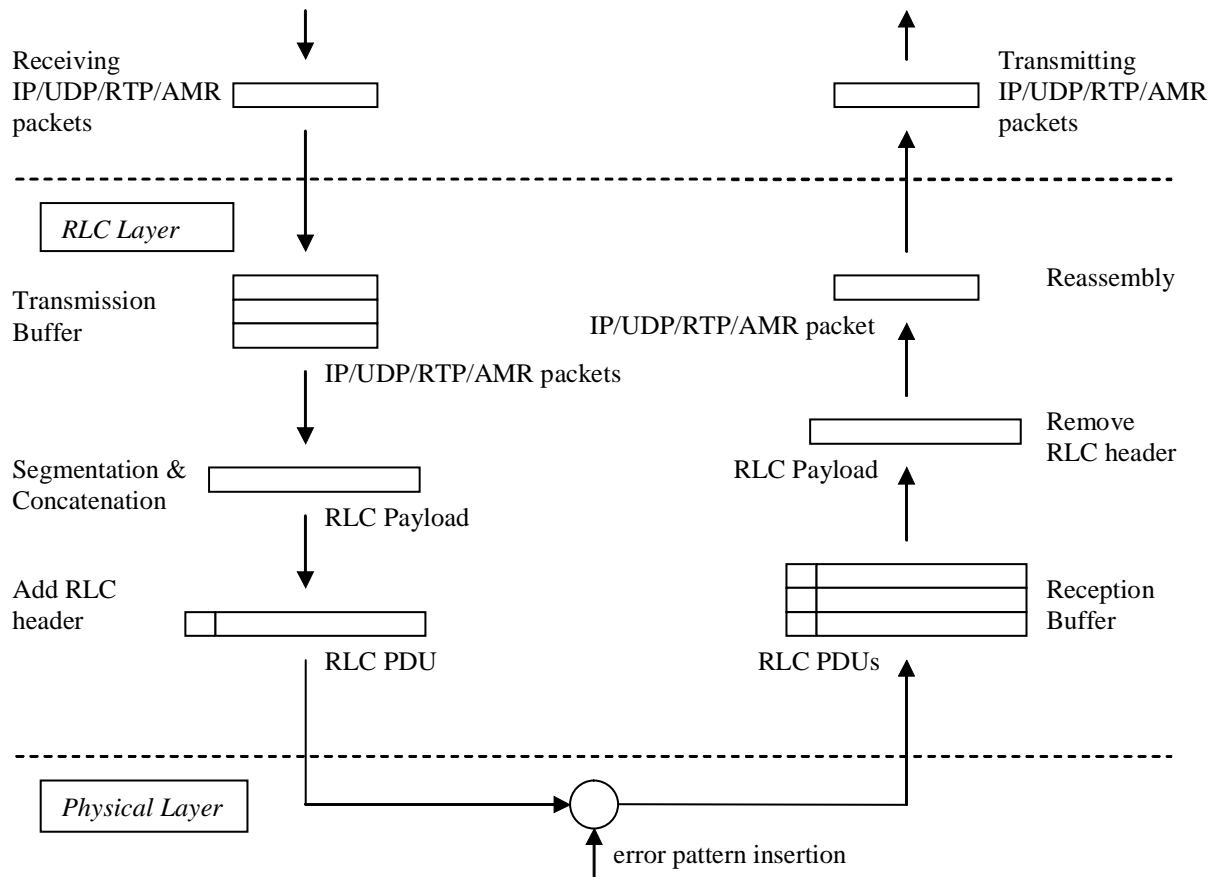
Higher layer	RAB/Signalling RB	<b>RAB</b>	
PDCP	PDCP header size, bit	8	
RLC	Logical channel type	DTCH	
	RLC mode	UM	
	Payload sizes, bit	920, 304, 96	
	Max data rate, bps	46000	
	UMD PDU header, bit	8	
MAC	MAC header, bit	0	
	MAC multiplexing	N/A	
Layer 1	TrCH type	DCH	
	TB sizes, bit	928, 312, 104	
	TFS	TF0, bits	0x928
		TF1, bits	1x104
		TF2, bits	1x312
		TF3, bits	1x928
	TTI, ms	20	
	Coding type	TC	
	CRC, bit	16	
	Max number of bits/TTI after channel coding	2844	
	Uplink: Max number of bits/radio frame before rate matching	1422	
	RM attribute	180-220	

### 3.2.3.2 Description of the RLC implementation

The UMTS air interface simulator (PC 2 and 4) receives IP/UDP/RTP/AMR packets on a specified port of the network card (see Figure 4). The IP/UDP/RTP/AMR packets are given to the transmission buffer of the RLC layer, which works in UM. The RLC will segment or concatenate the IP bitstream in RLC PDUs, adding appropriate RLC headers (sequence number and length indicators). It is assumed that always Transport Format TF 3 is chosen on the physical layer, providing an RLC PDU length including header of 928 bits. In the regular case, one IP packet is placed into an RLC PDU that is filled up with padding bits. Due to delayed packets from the network simulator it may also occur that there are more than one IP packets in the RLC transmission buffer to transmit in the current TTI.

Each TTI of 20ms, an RLC PDU is formed. It is then given to the error insertion block that decides if the RLC PDU is transmitted successfully over the air interface or if it is discarded due to a block error after channel decoding. The physical layer will not be simulated in real time, but error pattern files will be provided. The error patterns of the air interface transmission will be simulated according to the settings given in Section 0. They consist of binary decisions for each transmitted RLC PDU, resulting in a certain BLER.

After the error pattern insertion, the RLC of the air interface receiver site receives RLC PDUs in the reception buffer. The sequence numbers of the RLC headers are checked to detect when RLC PDUs have been discarded due to block errors. A discarded RLC PDU will result in one or more lost IP packets, resulting in a certain packet loss rate of the IP packets and thereby in a certain FER of the AMR frames. The IP/UDP/RTP/AMR packets are reassembled and transmitted to the next PC. This PC is either the network simulator (PC 3) in case of uplink transmission, or it is one of the terminals (PC 1 or 5) in case of downlink transmission.



**Figure 4:** UMTS air interface simulation

### 3.2.3.3 Physical Layer Implementation

The parameters of the physical layer simulation were set according to the parameters for a DCH in multipath fading conditions given in TS 34.121 (downlink) and TS 25.141 (uplink). The TB size is 928 bits and the Turbo decoder uses the Log-MAP algorithm with 4 iterations. The rake receiver has 6 fingers at 60 possible positions.

The different channel conditions given in Table 1, Table 2, and Table 3 were extracted from TR 101 112 (Selection procedures for the choice of radio transmission technologies of the UMTS) and also mentioned in the annex of the document S4-020680.

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	FLAT
2	50	-3.0	FLAT
3	110	-10.0	FLAT
4	170	-18.0	FLAT
5	290	-26.0	FLAT
6	310	-32.0	FLAT

**Table 1: Indoor Office Test Environment Tapped-Delay-Line Parameters**

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0.0	CLASSIC
2	310	-1.0	CLASSIC
3	710	-9.0	CLASSIC
4	1090	-10.0	CLASSIC
5	1730	-15.0	CLASSIC
6	2510	-20.0	CLASSIC

**Table 2: Vehicular Test Environment, High Antenna, Tapped-Delay-Line Parameters**

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	CLASSIC
2	110	-9.7	CLASSIC
3	190	-19.2	CLASSIC
4	410	-22.8	CLASSIC
5	-	-	CLASSIC
6	-	-	CLASSIC

**Table 3: Outdoor to Indoor and Pedestrian Test Environment Tapped-Delay-Line Parameters**



Table 4 (DL) and Table 5 (UL) show approximate results of the air interface simulation for  $\frac{DPCH - E_c}{I_{or}}$  and  $E_b/N_0$  corresponding to the considered BLERs.

Channel	BLER			
	$5*10^{-2}$	$1*10^{-2}$	$1*10^{-3}$	$5*10^{-4}$
Indoor, 3 km/h ( $\hat{I}_{or}/I_{oc} = 9$ dB)	-13.1 dB	-8.9 dB	-3.4 dB	-2.4 dB
Outdoor to Indoor, 3 km/h ( $\hat{I}_{or}/I_{oc} = 9$ dB)	-13.2 dB	-9.7 dB	-5.9 dB	-5.2 dB
Vehicular, 50 km/h ( $\hat{I}_{or}/I_{oc} = -3$ dB)	-9.35 dB	-8.2 dB	-6.9 dB	-6.55 dB
Vehicular, 120 km/h ( $\hat{I}_{or}/I_{oc} = -3$ dB)	-9.7 dB	-8.95 dB	-7.95 dB	-7.55 dB

**Table 4:** Downlink performance - approximately  $\frac{DPCH - E_c}{I_{or}}$  for the different channels and BLER

Channel	BLER			
	$5*10^{-2}$	$1*10^{-2}$	$1*10^{-3}$	$5*10^{-4}$
Indoor, 3 km/h	3.9 dB	6.4 dB	9.2 dB	9.8 dB
Outdoor to Indoor, 3 km/h	3.7 dB	6.1 dB	8.6 dB	9.2 dB
Vehicular, 50 km/h	-0.9 dB	-0.15 dB	0.55 dB	0.75 dB
Vehicular, 120 km/h	0.2 dB	0.6 dB	1.1 dB	1.3 dB

**Table 5:** Uplink performance - approximately  $E_b/N_0$  for the different channels and BLER

### 3.2.4 Headsets and Sound Card

To avoid echo problems, it has been decided to use headsets, instead of handsets. The monaural headsets are connected to the sound cards of the PCs supporting the AMR simulators.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, will not be modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid to place the acoustic opening of the microphone in front of the mouth.

### 3.2.5 Test environment

Each of the two subjects participating to the conversations is installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. When needed, the background noise is generated in the appropriate test room through a set of 4 loudspeakers. The background noise level is adjusted and controlled by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

### 3.2.6 Calibration and test conditions monitoring

#### Speech level

Before the beginning of a set of experiment, the end to end transmission level is checked subjectively, to ensure that there is no problem. If it is necessary to check the speech level following procedure will apply. An artificial mouth placed in front of the microphone of the Headset A, in the LRGP position -See ITU-T Rec. P.64-, generates in the artificial ear (according to ITU-T Rec. P57) coupled to the earphone of the

Head set B the nominal level defined in section 4.3. If necessary, the level is adjusted with the receiving volume control of the headset. The similar calibration is done by inverting headsets A and B.

### Delay

The overall delay (from the input of sound card A to the output of sound card B) will be evaluated for each test condition.

The hypothetical delay is calculated as shown :

On the air interface side, the simulator only receives packets on its network card, process them and transmits every 20 ms these packets to the following PC. Only processing delay and a possible delay due to a jitter can be added (a packet arrives just after the sending window of the air interface).

The hypothetical delay is calculated as shown :

On encoder side, delay have to take into account framing, look-ahead, processing and packetization: 45ms

Uplink delay between UE and Iu: 84.4 ms (see TR25.853)

Core network delay: a few ms

Routing through IP: depending on the number of routers.

Downlink delay between Iu and Ue: 71.8 ms (see TR25.853)

And delay on decoder side, taking into account jitter buffer, de-packetization and processing, 40 ms

The total delay to be considered is at least: 241.2 ms

### 3.3 Test Conditions

Based on circuit switched testing experiments, SA4 expects AMR 4.75 kb/s to provide insufficient quality for conversational applications. SA4 does not recommend testing AMR 4.75kb/s, this mode is considered as fall back solution in case of poor radio conditions.

Condition	Additional Background noise Room A	Additional Background noise Room B	Experimental actors		
			Radio conditions	IP conditions (Packet loss ratio)	Mode + delay
1	No	No	$10^{-2}$	0%	6,7kbit/s (delay 300 ms)
2	No	No	$10^{-2}$	0%	12,2 kbit/s (delay 500 ms)
3	No	No	$10^{-2}$	0%	12,2 kbit/s (delay 300 ms)
4	No	No	$10^{-2}$	3%	6,7kbit/s (delay 300 ms)
5	No	No	$10^{-2}$	3%	12.2kbit/s(delay 500 ms)
6	No	No	$10^{-2}$	3%	12,2 kbit/s (delay 300 ms)
7	No	No	$10^{-3}$	0%	6,7kbit/s (delay 300 ms)
8	No	No	$10^{-3}$	0%	12.2kbit/s(delay 500 ms)
9	No	No	$10^{-3}$	0%	12,2 kbit/s (delay 300 ms)
10	No	No	$10^{-3}$	3%	6,7kbit/s (delay 300 ms)
11	No	No	$10^{-3}$	3%	12.2kbit/s(delay 500 ms)
12	No	No	$10^{-3}$	3%	12,2 kbit/s (delay 300 ms)

13	No	No	$5 \cdot 10^{-4}$	0%	6,7kbit/s (delay 300 ms)
14	No	No	$5 \cdot 10^{-4}$	0%	12.2kbit/s(delay 500 ms)
15	No	No	$5 \cdot 10^{-4}$	0%	12,2 kbit/s (delay 300 ms)
16	No	No	$5 \cdot 10^{-4}$	3%	6,7kbit/s (delay 300 ms)
17	No	No	$5 \cdot 10^{-4}$	3%	12.6kbit/s(delay 500 ms)
18	No	No	$5 \cdot 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
19	Car	No	$5 \cdot 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
20	No	Car	$5 \cdot 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
21	Cafeteria	No	$5 \cdot 10^{-4}$	0%	6,7 kbit/s (delay 300 ms)
22	No	Cafeteria	$5 \cdot 10^{-4}$	0%	6,7 kbit/s (delay 300 ms)
23	Street	No	$5 \cdot 10^{-4}$	0%	12.2kbit/s(delay 500 ms)
24	No	Street	$5 \cdot 10^{-4}$	0%	12.2kbit/s(delay 500 ms)

#### Noise types

Noise type	Level (dB Pa A)
Car	60
Street	55
Babble	50

Listening Level	1	79 dB SPL
Listeners	32	Naïve Listeners
Groups	16	2 subjects/group
Rating Scales	5	
Languages	1	See table
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T, Recommendation P.800, Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1), except when background noise is needed (see table)

---

## Annex A Example Instructions for the conversation test

**Table :** Instructions to subjects.

<b>INSTRUCTIONS TO SUBJECTS</b>				
<p>In this experiment we are evaluating systems that might be used for telecommunication services. You are going to have a conversation with another user. The test situation is simulating communications between two mobile phones. The most of the situations will correspond to silent environment conditions, but some other will simulate more specific situations, as in a car, or in a railway station or in an office environment, when other people are discussing in the background.</p> <p>After the completion of each call conversation, you will have to give your opinions on the quality, by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgment will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the 5 following questions.</p>				
<p>Question 1: How do you judge the quality of the voice of your partner?</p>				
Excellent	Good	Fair	Poor	Bad
<p>Question 2: Do you have difficulties to understand some words?</p>				
All the time	Often	Some time to time	Rarely	Never
<p>Question 3: How did you judge the conversation when you interacted with your partner?</p>				
Excellent interactivity (similar to face-to-face situation)	Good interactivity (in few moments, you were talking simultaneously, and you had to interrupt yourself)	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt yourself)	Poor interactivity (often, you were talking simultaneously, and you had to interrupt yourself)	Bad interactivity (it was impossible to have an interactive conversation)
<p>Question 4: Did you perceive any impairment (noises, cuts,...)? In that case, was it:</p>				
No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment
<p>Question 5: How do you judge the global quality of the communication?</p>				
Excellent	Good	Fair	Poor	Bad
<p>From then on you will have a break approximately every 30 minutes. The test will last a total of approximately 60 minutes. Please do not discuss your opinions with other listeners participating in the experiment.</p>				

## Annex B: Example Scenarios for the conversation test

The pretexts used for conversation test are those developed by the Ruhr University (Bochum, Germany) within the context of ITU-T SG12 . These scenarios have been elaborated to allow a conversation well balanced within both participants and lasting approximately 2'30 or 3', and to stimulate the discussion between persons that know each other to facilitate the naturalness of the conversation. They are derived from typical situations of every day life: railways inquiries, rent a car or an apartment, etc. Each condition should be given a different scenario.

Examples coming from ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

- Scenario 1 : Pizza service

Subject 1:

Your Name :	Clemence
Reason for the call	1 large Pizza
Condition which should be applied to the exchange of information	For 2 people, Vegetarian pizza preferred
Information you want to receive from your partner	Topping Price
Information that your partner requires	Delivery address : 41 industry street,Oxford Phone : 7 34 20
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	How long will it take?

Subject 2:

Your Name :	Pizzeria Roma			
Information from which you should select the details which your partner requires	Pizzas	1 person	2 persons	4 persons
	Toscana (ham, mushrooms, tomatoes,cheese)	3.2£	5.95£	10.5£
	Tonno (Tuna, onions, tomatoes, cheese)	3.95£	7.5£	13.95£
	Fabrizio (salami, ham, tomatoes, heese)	4.2£	7.95£	14.95£
	Vegetaria (spinach, mushrooms, tomatoes,cheese)	4.5£	8.5£	15.95£
Information you want to receive from your partner	Name address telephone number			
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.				

- Scenario 2 : Information on flights

Subject 1:

Your Name :	Parker
Reason for the call	Intended journey: London Heathrow → Düsseldorf
Condition which should be applied to the exchange of information	On June 23th, Morning flight, Direct flight preferred
Information you want to receive from your partner	Departure : Arrival Flight number
Information that your partner requires	Reservation : 1 seat, Economy class Address: 66 middle street, Sheffield Phone: 21 08 33
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	From which airport is it easier to get into Cologne center : Düsseldorf or Cologne/Bonn

Subject 2:

Your Name :	Heathrow flight information			
Information from which you should select the details which your partner requires	Flight schedule	Lufthansa	British Airways	Lufthansa
	Flight number	LH 2615	BA 381	LH 413
	London Heathrow departure	6:30	6:35	8:20
	Brussels arrival		7:35	
	Brussels departure		8:00	
	Düsseldorf arrival	7:35	9:05	9:25
Information you want to receive from your partner	Name address telephone number number of seats Class : Business or Economy			
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.				

ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

### Annex C: Results to be provided

For contractual purposes, the information which needs to be provided is defined here.

The information required from each test Laboratory is a table containing the following information for each of the conditions in the experiment:

The "Mean Opinion Score (MOS)" obtained for all the subjects.

When the conditions are symmetrical, the mean value is calculated from all the result for the two test rooms..

For the dissymmetric conditions, the mean is calculated on the two test conditions, each result cumulating the results obtained in each condition of background noise.

The Standard Deviation of the "MOS" obtained for all the subjects, for each test condition.

The specific statistical comparisons are specified in Annex C.

## Annex D: Data analysis and presentation of results

### D.1 Calculation of MOS and Standard Deviation

The (overall) MOS/DMOS for confounded subjects for condition C ( $Y_c$ ) can then be obtained from:

$$Y_c = \frac{1}{T} \sum_{t=1}^T Y_{c,t}$$

The standard deviation (S) for condition C, denoted as  $S_c$ , can be calculated as:

$$S_c = \sqrt{\frac{1}{L \times T - 1} \sum_{t=1}^T \sum_{l=1}^L (X_{c,l,t} - Y_c)^2}$$

Finally, the confidence interval (CI) at the  $(1-\alpha)$  level can be calculated for  $N = L \times T$  as:

$$CI_c = (t_{1-\alpha, N-1}) \frac{S_c}{\sqrt{N}}$$

### D.2 Presentation of Basic Statistical Results

The test results should be reported by the test Laboratory and the Global Analysis Laboratory as follows: Calculate and tabulate "Mean Opinion Scores" for the (opinion scales, Standard Deviations and Confidence Intervals as shown in Table E.1.

Table C.1 - Layout for presentation of test results.

### D.3 Thorough analysis

Two statistical analyses should be conducted on the data obtained with these subjective scales. The first analysis consists in a Multiple ANalysis OF VAriance (MANOVA), which globally indicates the possible effect of the experimental factors (*i.e.*, different conditions). Then, a specific ANOVA should be run on each dependent variable (the five scales) to test if there is an effect of a specific experimental factor for a given subjective variable. In other words, these statistical analyses indicate if the differences observed between the MOS obtained for the different conditions are significant, for one given dependant variable (ANOVA) or for the whole of dependant variables (MANOVA). Finally, Pearson's linear correlations should be computed between the results of all subjective variables, to see which are those preponderant or dependent on others.

---