Technical Specification Group Services and System Aspects **TSGS#22(03)0679**
Meeting #22, Maui, Hawaii, USA, 15-18 December 2003 **Agenda Item: 7.4.3**

3GPP TSG-SA4 Meeting #29 *Tdoc S4-030747*
Tampere, November 24 – 28.

Source: **TSG SA WG4** (France Telecom R&D)
Title: **Test plan for 3G packet switched conversation tests - Phase 2: Comparison of quality offered by different speech coders**.

Document For: **Approval**
Agenda Item: **7.4.3**

1 Introduction

This document proposes a conversation test plan to compare the quality obtained with several different speech coders, over packet switched networks.
The different speech coders used in this test are
Adaptive Multi-Rate Narrow-Band (AMR-NB), in modes 6.7 kbit/s and 12.2 kbit/s,
Adaptive Multi-Rate Wide-Band (AMR-WB), in modes 12.65 kbit/s and 15.85 kbit/s,
ITU-T G.723.1, in mode 6.4 kbit/s,
ITU-T G.729, in mode 8 kbit/s,
ITU-T G.722, in mode 64 kbit/s, with packet loss concealment and,
ITU-T G.711, with packet loss concealment.
As there is no standardized packet loss concealment, plc for G.711 and G.722 are proprietary algorithms.

The simulated network will include two values of IP packet loss.

The test will be done in one test laboratory, only, but in two different languages.

This discussion gives references, conventions and contacts, section 3 details the test methodology, including test arrangement and test procedure,

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.

2. References, Conventions, and Contacts

2.1Permanent 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
	ITU-T Rec. G.711	Pulse code modulation (PCM) of voice frequencies
	ITU-T Rec. G.729	Coding of speech at8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)
	ITU-T Rec. G.723.1	Speech coders : Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s
	ITU-T Rec. G.722	7 kHz audio-coding within 64 kbit/s

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	L. Gros Laeticia.gros@francetelecom.com	France Telecom R&D	2, Avenue P. Marzin, 22307 Lannion Cédex France	Tel : +3329605 0720 Fax : +33296051316
AOB	Paolo Usai paolo.usai@etsi.fr	ETSI MCC	650 Route des Lucioles 06921 Sophia Antipolis Cedex 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**
1	France Telecom R&D	French
	France Telecom R&D	Arabic

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 networks simulators and communications are impaired by means of an IP impairments simulator part of the CN simulator, as the figure below describes it.

3.2 Test arrangement

3.2.1 Description of the proposed testing system

This contribution describes a networks simulator for the characterization of the different 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:

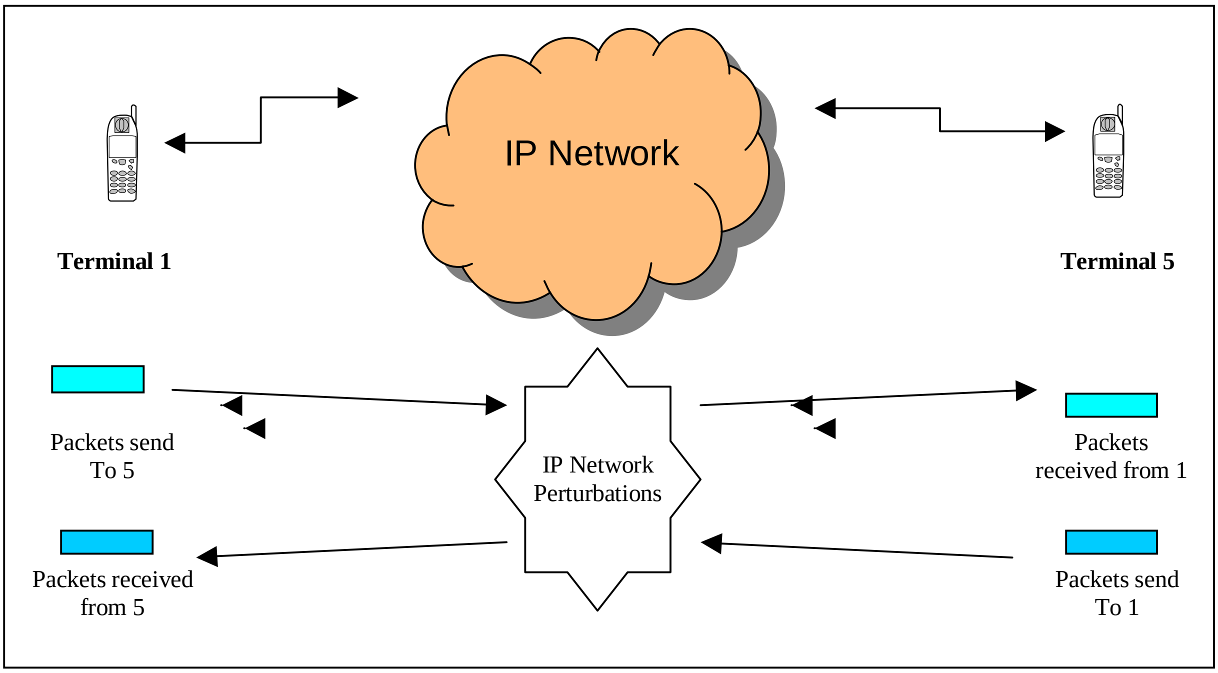


Figure 1: Packet switch audio communication simulator

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

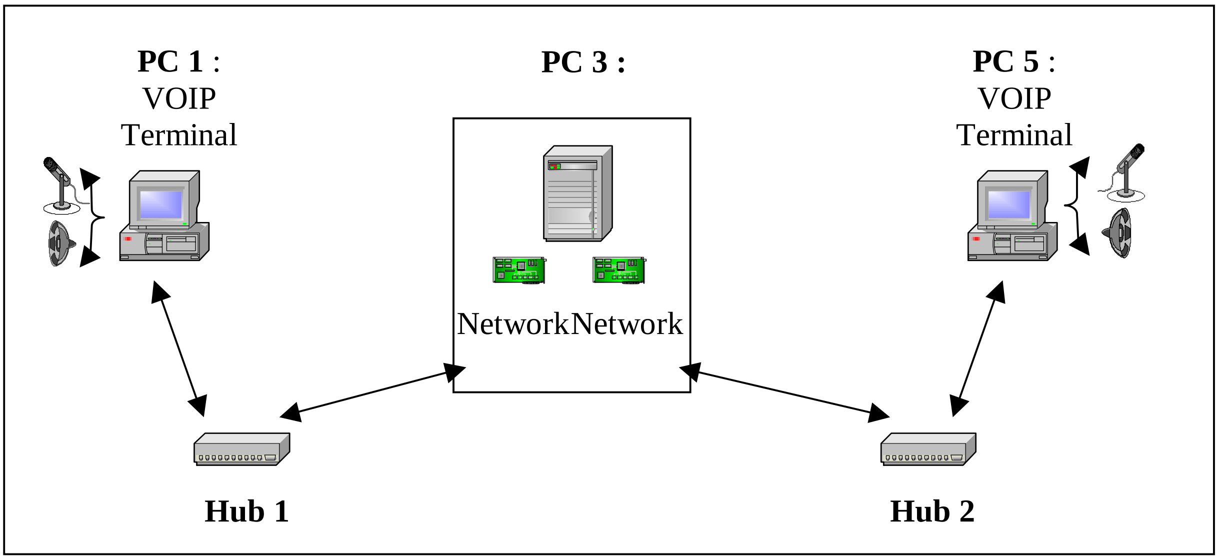


Figure 2: Simulation Platform

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

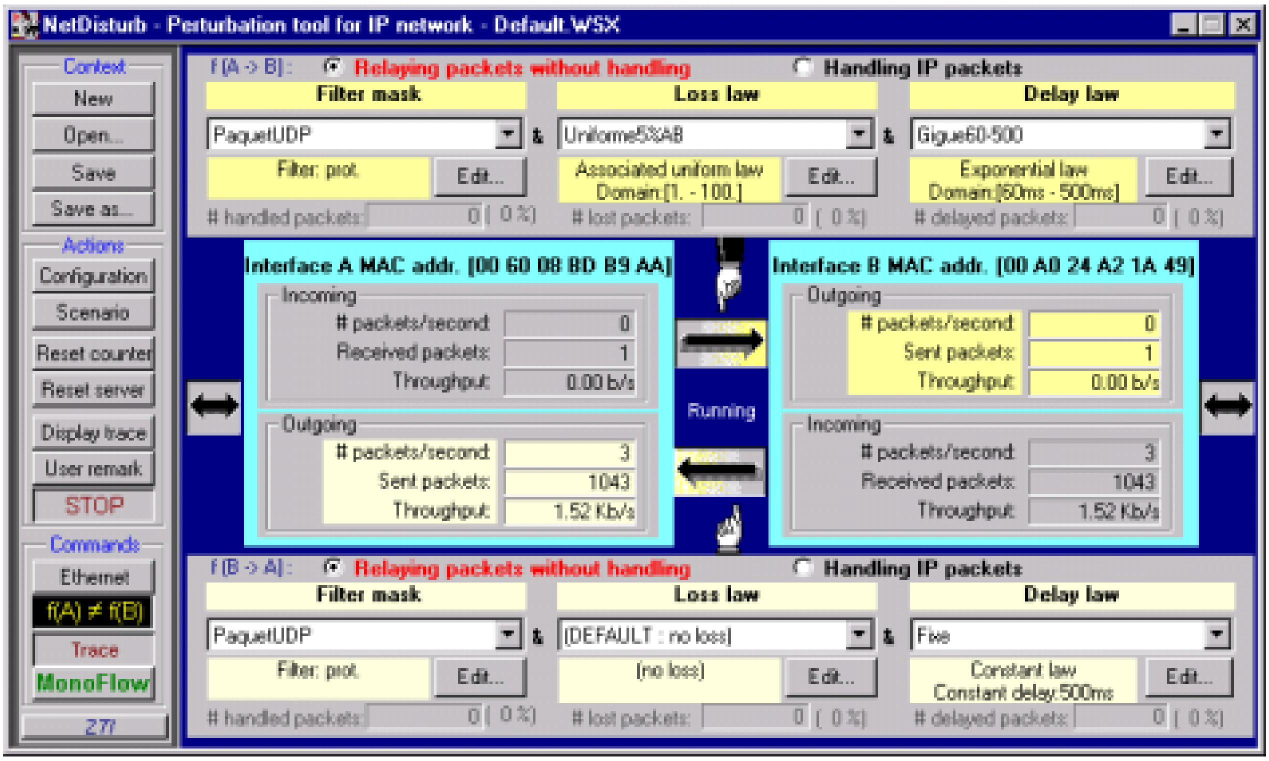
PC 3: PCs under WinNT OS with Network Simulator Software (NetDisturb).

Basic Principles:

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

3.2.2 France Telecom Network simulator

The core network simulator, as implemented, works under IPv4.
Figure 3 shows the possible parameters that can be modified, but, in this test, only "loss Law" will have two values, all the others settings being fixed.



NetDisturb - Perturbation tool for IP network - Default.WSX

f [A -> B] : ● Relaying packets without handling ○ Handling IP packets

Filter mask	Loss law	Delay law
PaquetUDP	Uniforme5%AB	Gigue60-500
Filter: prot.	Associated uniform law Domain:[1. - 100.]	Exponential law Domain:[60ms - 500ms]
# handled packets: 0 (0 %)	# lost packets: 0 (0 %)	# delayed packets: 0 (0 %)

Interface A MAC addr. [00 60 08 BD B9 AA]

Incoming

# packets/second	0
Received packets	1
Throughput	0.00 b/s

Outgoing

# packets/second	3
Sent packets	1043
Throughput	1.52 Kb/s

Running

Interface B MAC addr. [00 A0 24 A2 1A 49]

Outgoing

# packets/second	0
Sent packets	1
Throughput	0.00 b/s

Incoming

# packets/second	3
Received packets	1043
Throughput	1.52 Kb/s

f [B -> A] : ● Relaying packets without handling ○ Handling IP packets

Filter mask	Loss law	Delay law
PaquetUDP	[DEFAULT : no loss]	Fixe
Filter: prot.	(no loss)	Constant law Constant delay:500ms
# handled packets: 0 (0 %)	# lost packets: 0 (0 %)	# delayed packets: 0 (0 %)

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 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.

3.2.4 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. The background noise level is checked 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.

3.2.5 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 adjusted for each test condition taking into account the delay of the related codec in order to have a fixed delay around 250ms. This value of 250ms is close to the hypothetical delay computed for AMR and AMRWB through the UMTS network.

3.3 Test Conditions

Condition	**Experimental actors**		
		IP conditions (Packet loss ratio)	Mode
1		0%	AMR NB 6,7kbit/s
2		0%	AMR-NB 12,2 kbit/s
3		0%	AMR-WB 12,65 kbit/s
4		0%	AMR-WB 15,85 kbit/s
5		0%	G. 723.1 6,4 kbit/s
6		0%	G.729 8 kbit/s
7		0%	G.722 64 kbit/s + plc
8		0%	G.711 + plc
9		3%	AMR NB 6,7kbit/s
10		3%	AMR-NB 12,2 kbit/s (delay 300 ms)
11		3%	AMR-WB 12,65 kbit/s
12		3%	AMR-WB 15,85 kbit/s
13		3%	G. 723.1 6,4 kbit/s
14		3%	G.729 8 kbit/s
15		3%	G.722 64 kbit/s + plc
16		3%	G.711 + plc

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners per language
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),

1 References

Tdoc S4-030564- **Test Plan for the AMR Narrow-Band Packet switched Conversation test**

Tdoc S4-030565- **Test Plan for the AMR Wide-Band Packet switched Conversation test**

END

Annex A Example Instructions for the conversation test

Table : Instructions to subjects.

INSTRUCTIONS TO SUBJECTS

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. All the situations will correspond to silent environment condition
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.

Question 1: How do you judge the quality of the voice of your partner?

Excellent	Good	Fair	Poor	Bad

Question 2: Do you have difficulties to understand some words?

All the time	Often	Some time to time	Rarely	Never

Question 3: How did you judge the conversation when you interacted with your partner?

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)

Question 4: Did you perceive any impairment (noises, cuts,…)? In that case, was it:

No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment

Question 5: How do you judge the global quality of the communication?

Excellent	Good	Fair	Poor	Bad

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.

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:

<table>
<tr><td>Your Name :</td><td>Pizzeria Roma</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><th>Pizzas</th><th>1 person</th><th>2 persons</th><th>4 persons</th></tr>
<tr><td>Toscana
(ham, mushrooms, tomatoes, cheese)</td><td>3.2£</td><td>5.95£</td><td>10.5£</td></tr>
<tr><td>Tonno
(Tuna, onions, tomatoes, cheese)</td><td>3.95£</td><td>7.5£</td><td>13.95£</td></tr>
<tr><td>Fabrizio
(salami, ham, tomatoes, cheese)</td><td>4.2£</td><td>7.95£</td><td>14.95£</td></tr>
<tr><td>Vegetarian
(spinach, mushrooms, tomatoes, cheese)</td><td>4.5£</td><td>8.5£</td><td>15.95£</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number</td></tr>
<tr><td>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.</td><td></td></tr>
</table>

- 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 23rd, 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:

<table>
<tr><td>Your Name :</td><td>Heathrow flight information</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Flight schedule</td><td>Lufthansa</td><td>British Airways</td><td>Lufthansa</td></tr>
<tr><td>Flight number</td><td>LH 2615</td><td>BA 381</td><td>LH 413</td></tr>
<tr><td>London Heathrow departure</td><td>6:30</td><td>6:35</td><td>8:20</td></tr>
<tr><td>Brussels arrival
Brussels departure</td><td></td><td>7:35
8:00</td><td></td></tr>
<tr><td>Düsseldorf arrival</td><td>7:35</td><td>9:05</td><td>9:25</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number
number of seats
Class : Business or Economy</td></tr>
<tr><td>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.</td><td></td></tr>
</table>

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 (Yc) 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 Sc, 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-α) 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.