3GPP TSG RAN Meeting #19 RP-030127
Birmingham, United Kingdom, 11 – 14 March 2003

3GPP TSG-SA 4 Meeting #25bis **S4-030260**
Berlin, Germany, 24th-28 February .

Title: LS on Radio Access Bearer for PS conversational testing

Release: Rel. 5

Work Item: Performance characterisation of default codecs for PS conversational multimedia application (SA4)

Source: SA 4

To: RAN 2, TSG RAN, TSG CN

Cc: GERAN 2

Contact Person:
Name: Hervé Taddei
Tel. Number: +49 89 722 59068
E-mail Address: herve.taddei@siemens.com

Attachments: S4-030165

1. Overall Description:

At SA4#25 bis (24-28 Feb'03, Berlin), there was a proposal from France Telecom to set up a conversational test on PS conversational services. The purpose of this conversational test is to characterize the AMR and AMR WB used in PS voice service. To progress a conversation test methodology has been proposed, but still some test parameters (delay, packet loss, radio condition) are unsure.

The transmission of the voice is done using IP/UDP/RTP protocols with one AMR (or AMR-WB) frame per IP packet (every 20 ms when VAD is not used). Assuming IPv4 the size of the IP packets including the AMR will range from 53 to 71 bytes or 57 to 100 bytes for the AMR WB. The use case description can be found in TS. 26.236 Annex B use case 1.

In document TS 34.108 (Annex B paragraph 6.10.2.4.1.59.1.1.1 Transport channel parameters for Conversational / speech / UL:42.8 kbps / PS RAB) a RAB is proposed for PS conversational case.

2. Actions:

To RAN and CN groups.

ACTION:

SA 4 kindly asks the groups to answer our following questions:

- Is this example RAB the only one available for that type of service?
- If the previous statement is not right, could you provide us with the right and most suitable RAB parameters knowing the service we want to set (as described in the overall description)?
- Are the 100 ms transfer delay defined in the QoS (26.236 Use case 1) feasible on an UTRAN bearer (between the GGSN and the terminal)?
- Is it the understanding of RAN that the end to end delay is the sum of the 2 transfer delays plus the CN delay? Are there more delays to be taken into account?

3. Date of Next TSG-SA 4 Meetings:

TSG-SA 4 Meeting #26	5th – 9th May 2003	tbd.
TSG- SA 4 Meeting #27	7th–11th July 2003	Munich, Germany

3GPP TSG-S4 Meeting #25 bis ***Tdoc S4-030165***
Berlin, Germany, 24-28 february 2003

Source: **France Telecom**[1]

Title: **Test methodology for AMR NB Packet-switched conversational tests**

Document for: **Discussion**

Agenda Item: **7**

Summary

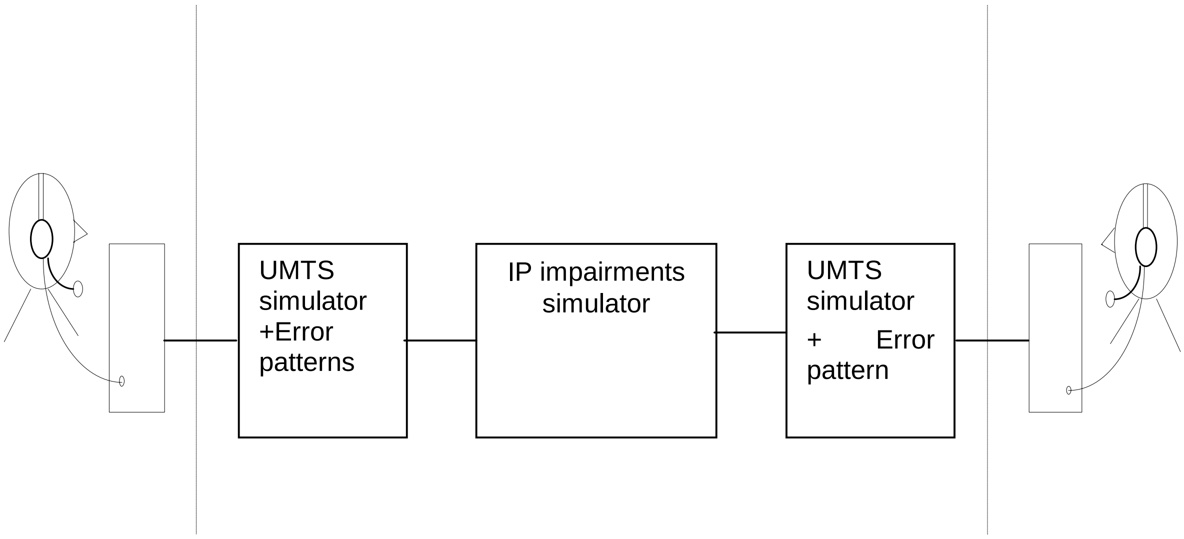
At previous meetings, it was decided to organize conversation tests for AMR and AMR-WB used in packet switch voice service. To progress, a conversation test methodology with some test parameters (delay, packet loss, radio conditions) is proposed for discussion. This methodology is based on ITU-T Recommendations.

Conversation test methodology

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).

Experimental setup and impairments of the quality of the speech transmission

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, as the figure below describes it.



The detailed test set-up is available in Annex 1
Each communication corresponds to one tested condition defined by a combination of the different factors exposed in the table below. All the combinations are performed, that gives eighteen conditions per experiment (3 radio conditions x 2 IP conditions x 3 AMR modes).

The eighteen conditions are presented with different modes of the AMR NB codecs, combined with Radio conditions and IP conditions.

[1] **Jean-Yves Monfort** Tel: +33 2 96 05 31 71 jeanyves.monfort@francetelecom.com
France Télécom R&D
2 Avenue P. Marzin, 22307 Lannion, France

Factor	Levels
Radio conditions	10^{-2} 10^{-3} $5\ 10^{-4}$
IP conditions	- Delay (50 ms), 0% of packet losses, 0 ms of jitter - Delay (50 ms), 3% of packet losses, 0 ms of jitter
Mode	4,75 kbit/s 6,7 kbit/s 12,2 kbit/s

Asymmetric conditions are included in the test plan. The dissymmetry is created to simulate different environmental noise conditions, simulating some usage scenarios

Subjects

Conversation-opinion tests are set up between non expert adults, which do not present auditory problem. A minimum of sixteen subjects is needed in order to conduct relevant statistical analysis of the data. However the larger the number of subjects, the easier the resolution of significant differences between different transmission conditions, as suggested in Tdoc S4-020684. A number of 32 subjects (*i.e.* 16 couples) seems to be a good choice.

Procedure

Each communication lasts between 2 or 3 minutes. This duration is sufficient for the subjects to make their opinion on quality. A longer duration of the communications entails a longer test in useless way. It has to be notice that a test can not last more than about two hours. Otherwise the test would become too tiring for the subjects, resulting in irrelevant judgments.

The pretexts used for this protocol are those developed by the Ruhr University (Bochum, Germany) within the context of ITU-T SG12 [5]. These scenarios have been elaborated to allow a conversation well balanced within both participants and lasting approximately 2'30 or 3'. These scenarios allow to stimulate the discussion between persons and 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.

After each communication (corresponding to one specific condition) the subjects have to judge the quality of the communication filling in a specific form. To be sure to assess all aspects of speech quality, five different questions are submitted to subjects after each of the communications they have when testing the terminals. These questions are extracted from the recommendations ITU-T P.800 [1], ITU-T P.830 [2], ITU-T P.831[3], ITU-T P.832 [4]. and allow to cover different aspects of quality (interactivity, quality of the interlocutor,…). Tdoc S4-020684 suggested seven-point scales. However it is important to keep the five-point scales generally recommended by the ITU-T. Therefore, it will be possible to compare results obtained with this test with previous results obtained for other networks, codecs, etc….The reference scale is the five-point MOS scale and it is really advisable to stay in this reference (ITU-T P.800). The five five-point category scales are the following;

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	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt	Poor interactivity (often, you were talking simultaneously, and you had to interrupt	Bad interactivity (it was impossible to have an interactive conversation)

	yourself)	yourself)	yourself)	

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

Remark: Tdoc S4-020684 recommends that the design of any questionnaire be based on general rather specific aspects of speech communication, for example, "Effect of System on communication" and "Effect on System on task performance" instead of "Perceived Impairments" (cf question 4) and "Perceived Echo". However, an analytic approach of the quality, according to criteria based on easy detectable and quantifiable aspects (such as impairments, echo, etc…), is easier for naïve subjects. Questions on the System itself are more intended to experts. Questions as they appeared above were validated in a previous conversation-opinion test on AMR in IP context. Subjects reported that these scales are well understandable and allowed them to easily express their opinion on quality.

Two statistical analyses are 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 is 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. Finally, correlations are computed between the results of all subjective variables, to see which are those preponderant or dependent on others.

Remark: The echo loss values as defined in TS131 are sufficient to guarantee an echo free transmission. Therefore, a first test should be carried out without considering echo condition. However, the conversation situation in which echo could appear should be only with a link to fixed network interconnected with a gateway or with IP phone, in the case of deficient signaling for echo cancellers. Therefore a second test should be performed in these echo condition, for AMR and AMR WB. The rating procedure should be identical that the one described above, with one added criterion:

Did you perceive any echo? In that case, was it?

No echo	Slight echo, but not disturbing	Echo slightly disturbing	Echo disturbing	Very disturbing echo

References

[1] Rec. ITU-T P.800, “Methods for subjective determination of transmission quality”, Geneva 1996.
[2] Rec. ITU-T P.830, “Subjective performance assessment of telephone-band and wideband digital codecs”, Geneva 1996.
[3] Rec. ITU-T P.831
[4] Rec. ITU-T P.832
[5] ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

ANNEX 1: Test Set-up

UMTS Simulator

The UMTS simulator is constituted of an AMR codec operating in floating point version.. The floating point version of AMR is running in real time on a PC. The sound pick-up and reproduction are done via sound cards included in the PCs.
The conditions of usage are taken from the existing 3GPP specifications, The payload is constituted following the IETF RFC3267 and 3GPP TS 26.235. Among the different offered options, to reflect the decision of 3GPP in the TS 26.236, the following ones have been chosen:
- the bandwidth efficient mode is used,
- only one speech frame is encapsulated in each RTP packet,
  - the multi-channel session is not used,
  - interleaving is not used,
- internal CRC is not used.

The UDP/IP encapsulation is done by the PC and the packet are delivered to the IP impairment system via an Ethernet interface.
The packet size depends on the mode of AMR as shown in the table below:

AMR mode	Bits/frame	Bytes/packet (assuming IPv4)
4.75	95	53
6.7	134	58
12.2	244	71

Error Pattern

SeeTdoc S4 –030035 From Siemens

Internet Simulator

The Radcom Internet Simulator is designed to monitor 5 parameters which are inherent to IP transmission. The parameters are the packet loss (Loss), bandwidth limitation (link), IP packet order (order), the transmission delay (latency) and the jitter.
The Internet simulator is a PC including two LAN 10/100 Mbps cards. The software generating the IP transmission impairments is able to generate independently the 5 parameters for each direction path. However for these conversation tests, only the parameter "Packet loss" is variable and for each test condition, the settings are similar for the two transmission paths

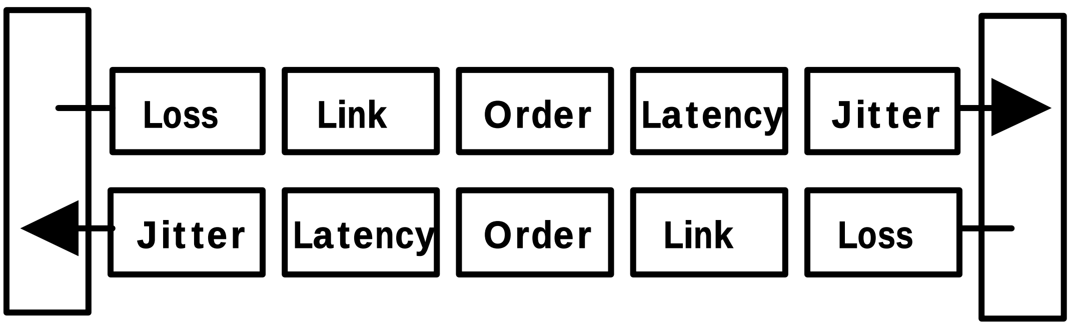


Figure 2 : Internet Simulator

For the experiment, the parameters are set as followed

No order effect
No bandwidth limitation
No transmission delay
No jitter
Packet loss set to 0%, 3%.

The Packet loss, simulated by the "Packet loss Simulator", is defined by using a Markovian law.

Headsets and Sound Card

To avoid echo problems, it has been decided to use headsets, instead of handsets. The binaural headsets are connected to the sound cards of the PCs supporting the AMR simulators.
The sound level in the earphones could be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, were not modified.
The microphones where protected by a foam ball to reduce the "pop" effect. It was also suggested to the user to avoid to place the acoustic opening of the microphone in front of the mouth.