Joint TSG-S4#10 - SMG11#15 Meeting **S4/SMG11 Tdoc 130-/00**
Helsinki, Finland
February 28th - March 3rd 2000

Title: Proposal for AMR-NS test plan (CCR)

Source: Motorola

General Philosophy:

- Reduce total number of conditions to 48 (24 of A/B comparison, 24 of B/A comparison) per experiment
- Reduce from 4 to 2 experiments by combining 'SNR' and 'SNR+9' experiments
- There is not suffficient 'room' to repeat all the ideal noise reference conditions for SNR+9. Therefore it is assumed that the main SNR improvement measurement will occur for the 'SNR' subset of conditions and that for 'SNR+9':
  - Improvement for car and street only will be measured
  - Only 2 ideal suppression levels will be tested, and these are relatively low (3 and 6dB) since very large improvement is not expected when starting with conditions which are already high SNR.
- Additionally, the Ideal noise suppression reference 'values' are shifted to 3, 6, 9dB to help judge against a proposed minimum SNR improvement requirement.

Note that section x.12 has been placed at the front, since it is the major change to be discussed and agreed.

x.12 Test Conditions

Cond.	Noise	SNR (dB)	Reference	Processed		Speech sample number (6 sequences)
				Ideal NS	Codec	
1	Car	6	AMR@x	-	AMR@x	4 5 6 1 2 3
2	Street	9	AMR@x	-	AMR@x	4 5 6 1 2 3
3	Babble	9	AMR@x	-	AMR@x	4 5 6 1 2 3
4	Car	6	MNRU-16	-	MNRU-12	4 - - 1 - -
5	Car	6	Direct	-	MNRU-12	4 - - 1 - -
4'	Street	9	MNRU-16	-	MNRU-12	- 5 - - 2 -
5'	Street	9	Direct	-	MNRU-12	- 5 - - 2 -
4"	Babble	9	MNRU-16	-	MNRU-12	- - 6 - - 3
5"	Babble	9	Direct	-	MNRU-12	- - 6 - - 3
6	Car	6	AMR@x	3	AMR@x	1 2 3 4 5 6
7	Car	6	AMR@x	6	AMR@x	1 2 3 4 5 6
8	Car	6	AMR@x	9	AMR@x	1 2 3 4 5 6
9	Street	9	AMR@x	3	AMR@x	2 3 4 5 6 1
10	Street	9	AMR@x	6	AMR@x	2 3 4 5 6 1
11	Street	9	AMR@x	9	AMR@x	2 3 4 5 6 1
12	Babble	9	AMR@x	3	AMR@x	3 4 5 6 1 2
13	Babble	9	AMR@x	6	AMR@x	3 4 5 6 1 2
14	Babble	9	AMR@x	9	AMR@x	3 4 5 6 1 2
15	Car	6	AMR@x	-	AMR/NS1@x	1 2 3 4 5 6
16	Street	9	AMR@x	-	AMR/NS1@x	2 3 4 5 6 1
17	Babble	9	AMR@x	-	AMR/NS1@x	3 4 5 6 1 2
18	Car	15	AMR@x	6	AMR@x	1 2 3 4 5 6
19	Car	15	AMR@x	9	AMR@x	1 2 3 4 5 6
20	Street	18	AMR@x	6	AMR@x	2 3 4 5 6 1
21	Street	18	AMR@x	9	AMR@x	2 3 4 5 6 1
22	Car	15	AMR@x	-	AMR/NS1@x	1 2 3 4 5 6
23	Street	18	AMR@x	-	AMR/NS1@x	2 3 4 5 6 1
24	Babble	18	AMR@x	-	AMR/NS1@x	3 4 5 6 1 2
25-48	Reversed order of the reference and processed speech samples in cond. 1-24					
Notes:	- AMR@x denotes AMR at bit rate x, AMR/Nsy@x denotes NS candidate y at bit rate x; 5.9 kbit/s for Experiment 4, 12.2 kbit/s for Experiment 5 - SNR(dB) denotes SNR for noise - 4 talkers are used for all conditions: 2 male and 2 female - 6 speech samples (8 s) are used for each talker					

x. Experiments 4a, 4b, 5a and 5b: Performances in Background Noise Conditions (CCR)

x.1 Introduction

These experiments are designed to test Requirement [TBD] in the associated Section in the Recommended Minimum Performance Requirements Specification (TS GSM [TBD]). Specifically, the AMR with noise suppression should be preferred to the AMR without noise suppression in a background noise environment and should provide a reasonable level of SNR improvement. Experiment 4 examines the performance of the noise suppression with the half-rate codec, while Experiment 5 examines the noise suppression with the full rate codec. Both experiments will use the Comparison Category Rating (CCR) method with a seven-point rating scale. Listeners will judge the relative quality of samples processed through the codec with noise suppression, compared to those without the noise suppression applied (instructions for listeners are given in Annex [TBD]). The samples will have background noise of various types and levels mixed into the source speech before processing through the codec.

The factors for each of the four sub-experiments are presented in Table x.1.1.

Factor	*Expt 4*	*Expt 5*
codec	AMR 5.9 kb/s	AMR 12.2 kb/s
noise types	car (6 and 15 dB) street (9 and 18 dB) babble (9 and 18 dB)	car (6 and 15 dB) street (9 and 18 dB) babble (9 and 18 dB)

Table x.1.1: Factors for Experiments 4 and 5

x.2 Test Factors and Conditions

Three types of background noise will be used, at two different SNRs:

- A car noise that is stationary both in level and in spectrum.
- A street noise that is non-stationary in level, but fairly stationary in spectrum.
- A babble noise that is fairly stationary in level, but non-stationary in spectrum.

The noise samples will be those utilised during the AMR Noise Suppression Selection Phase [1]
The codec is held constant for each experiment, with two SNR classes ('SNR' and 'SNR+9dB') per experiment. All of the noise types are used in each experiment. The noise samples will have been recorded in scenarios representative of the respective SNR value for each noise. (*Ed note: This presents a problem unless we can say that the recordings made at 'SNR+6dB' for the selection contest are also representative of 'SNR+9dB')*

The factors and conditions to be used in Experiments 4 and 5 are presented in Table x.2.1. The expanded set of test conditions is given in Section x.12.

Main Codec Conditions	**#**	**Notes**
Noise Suppresser Candidates	1	
Codec	1	AMR
Codec Modes (HR/FR)	HR FR	5.9 kbit/s rate for Experiment 4 12.2 kbps rate for Experiment 5
BERs	0	Clear channel, no transmission errors
Input level	1	nominal: -26dB relative to OVL
Acoustic Background Noise	3	car, street, and babble noise
Background noise SNRs	2	low, high for each (see Table x.2.1)
Input Characteristic	1	GSM transmit filtered
Codec references	**#**	**Notes**
All Experiments	1	the same AMR rate w/o NS
Other references	**#**	**Notes**
Direct		nominal level, GSM transmit filtered
MNRU, Exp 4 and 5		nominal level, GSM transmit filtered, Q= 12, ΔQ= 4
Ideal noise suppression simulation		
Common Conditions	**#**	**Notes**
GSM Channel	0	NO channel model
Number of talkers	4	2 male + 2 female primary talkers
Number of speech samples	28	7 Sentence-pairs/primary talker (6 for Test, 1 for Practice)
Listening Level	1	-15dBPa (79dB SPL) at ERP
Listeners	24	Naive Listeners
Randomizations	6	6 groups of 4 listeners
Rating Scale	1	CCR Instructions
Replications	1	Original Presentation Only

Table x.2.1: Factors and conditions for Experiments 4 and 5

x.3 Preliminary Conditions

The following 16 preliminary test conditions are recommended, for presentation, before proceeding to the test samples. The samples shall be presented in the random order given in Table x.3.1

(Ed Note - I haven't altered these, but they need to change to allow for 'SNR+9dB' in the same expt as 'SNR' - do we simply double the number of preliminaries?)

Cond.	**Presentation order**	Noise	SNR (dB)	Reference	Processed Ideal NS	Processed Codec	Talker and Sample Number
P1	9	Car	SNR_C	Direct	-	Direct	M1S07
P2	5	Car	SNR_C	AMR@x	-	AMR@x	F1S07
P3	12	Car	SNR_C	MNRU-12	-	MNRU-16	M2S07
P4	13	Car	SNR_C	MNRU-12	-	Direct	F2S07
P5	2	Street	SNR_S	AMR@x	-	AMR@x	M1S07
P6	4	Street	SNR_S	MNRU-12	-	MNRU-16	F1S07
P7	8	Street	SNR_S	MNRU-12	-	Direct	M2S07
P8	16	Babble	SNR_B	AMR@x	-	AMR@x	F2S07
P9	7	Babble	SNR_B	MNRU-12	-	MNRU-16	M1S07
P10	1	Babble	SNR_B	MNRU-12	-	Direct	F1S07
P11	11	Car	SNR_C	AMR@x	4	AMR@x	M2S07
P12	3	Car	SNR_C	AMR@x	10	AMR@x	F2S07
P13	15	Street	SNR_S	AMR@x	4	AMR@x	M1S07
P14	6	Street	SNR_S	AMR@x	10	AMR@x	F1S07
P15	10	Babble	SNR_B	AMR@x	4	AMR@x	M2S07
P16	14	Babble	SNR_B	AMR@x	10	AMR@x	F2S07
Notes:	- The bit rate for the AMR processing for the preliminary samples shall be the same as that used for the test samples, 5.9 kbit/s for Experiment 4, 12.2 kbit/s for Experiment 5. - SNR_C/S/B denotes SNR for noise [TBD]						

Table x.3.1: List of preliminary conditions

x.4 Speech Material

The source speech material shall be as defined in Section [TBD] and will consist of the material used during the AMR Noise Suppression Selection phase: Each sample consists of two sentences. Only primary talkers are needed. For the four talkers, the following source material should be prepared:

> Seven samples for each talker, six for the test samples and one for the preliminaries,
> Each sample to be eight seconds long,
> Unique sentences-pairs in each sample (i.e., no repeated across the talkers)

To reduce any speech material effect, the samples for each talker must be unique. For these experiments, these unique stimuli are not balanced across all conditions, candidates and subject groups. The same sample numbers for each talker are used for common conditions within a subject group and changed across subject groups (these sample numbers are arbitrarily assigned to samples). For a given language, the same speech material must be used for the two experiments 4 and 5.The noise material and its mix with the speech material should be as defined in Section [TBD] and Section [TBD].

x.5 Experimental Design

The design is based on a restricted randomization philosophy using six different randomizations, each of which is used with a group of four of the 24 listeners. This means that up to four subjects can perform the experiment simultaneously.

Each listener will hear all of the conditions four times, once with speech from each of the four talkers. Over the experiment as a whole, each of the conditions will be paired with six different samples from each of the four talkers. Each of the six groups of subjects will hear different combinations of source material and condition.

x.6 Processing

Every condition is processed with each of the six samples of each of the four primary talkers. The actual samples to be used for each condition, within with each subject group, are presented in Section x.12, *Test Conditions.*

x.7 Randomizations

The test shall be completed using the randomizations provided in Annex [TBD]. There are six randomizations for the sub-experiments, one for each subject group. The same randomizations will be used for the two experiments (4 and 5). Each one will therefore be used by four of the 24 subjects. Each randomization is balanced across four blocks of 48 stimuli to eliminate long sequences of similar conditions or identical talkers. The sequences provide for alternating male-female talkers. Use of these randomizations will allow presentation order to be used as a factor in a global analysis, should that be necessary. The randomization is constrained to a randomized block design, which controls practice and fatigue effects that may occur over the course of a test session.

x.8 Duration of the DCR Experiments 4 and 5

Each trial consists of an eight-second reference sample + an eight-second test sample + five second voting time, totaling 21 seconds. For each of the four experiments there are [TBD] preliminary conditions x 21 seconds or [TBD] minutes for an introductory block. Each presentation set within an experiment consists of 48 conditions (A/B+B/A) x 4 talkers x 21 seconds or 67.2 minutes, presented as eight 8.4 minute blocks of 32 stimuli for [TBD] minutes testing time / subject group / experiment. The total testing time for each experiment will be [TBD] hours and [TBD] minutes, if four listeners are tested at one time.

To reduce the effects of subject fatigue, each 8.4 minute block should be separated by short comfort breaks.

Note that the above calculations do not include the time needed to give the subjects their instructions, or time taken for comfort breaks.

x.9 Votes Per Condition

In each of the three experiments, 24 listeners rate every condition with four talkers in each of two presentation orders (A/B and B/A), giving:

(24 subjects x 4 talkers x 2 presentations) = 192 votes per condition

From past experience with CCR tests, this is the minimum number of votes per condition needed to give enough statistical certainty to differentiate the performance of one candidate process from another candidate process over the conditions and against the references.

x.10 Test Procedure

Factors important for the experimental environment are specified in Sections [TBD] and [TBD]. As specified in Section x.8, comfort breaks should be provided to reduce the effects of subject fatigue.

x.11 Opinion Scale

The question asked of the subject is a based on of the CCR Listening Quality Comparison Scale. The listening subjects will judge the quality of the second sample with regard to quality of the first sample. The subjects will listen to each pair of samples and after these have been played, they will be asked to give their comparative opinion. Annex [TBD] contains an example of the instructions for the subjects in English. Changes to the instructions may be needed to specify the method of data collection being used (button-press, paper & pencil, etc.).

References

1 Test plan v1.9