Technical Specification Group Services and System Aspects *TSGS#28 (05)0254*
Meeting #28, Quebec, Canada, 6-9 June 2005

Source: **TSG-SA WG4**

Title: PSS/MMS/MBMS Audio Codec Characterization Test Plan (ext. to Phase 2)

Agenda Item: **7.4.3**

Presentation to: **TSG SA Meeting #28**

Presented for: **Approval**

Abstract of document:

The present document contains the **PSS/MMS/MBMS Audio Codec Characterization Test Plan** (extension to Phase 2).

The Performance Characterisation tests (Phase 2) for the *Extended Adaptive Multi-Rate - Wideband* (AMR-WB+) and the *Enhanced aacPlus* codecs will be performed by Fraunhofer Institut, Nokia, NTT_AT, and T-Systems.

The Global Analysis of the Performance Characterisation tests (Phase 2) for the *Extended Adaptive Multi-Rate - Wideband* (AMR-WB+) and the *Enhanced aacPlus* codecs will be performed by Dynastat.

The results of Phase 2 of testing will be included in TR 26.936 "*Performance characterization of the Enhanced aacPlus and Extended Adaptive Multi-Rate - Wideband (AMR-WB+) audio codecs (Release 6)*", c/o Dynastat.

The approval of the **PSS/MMS/MBMS Audio Codec Characterization Test Plan (ext. to Phase 2)** is requested to SA#28 Plenary for ETSI to proceed to contract the laboratories mentioned above.

Outstanding Issues:

None.

Contentious Issues:

None.

3GPP TSG-SA4 Meeting #35
San Diego, California, 9-13 May 2005

Tdoc S4-050440

Title: **PSS/MMS/MBMS Audio Codec Characterization Test Plan**

Source: **Editor [Dynastat]**

Agenda Item: **7.4.3**

Contents

0 Revision History

Ver.	Meeting	Location	Doc.	Changes
0.1	SA4#32	Prague	S4-040538	Initial draft of the test plan
0.2	SA4#33	Helsinki	S4-040790	Organization of tests
0.3	AC-AdHoc	Munich	AHAUC-026	Drafting of Phase 1 tests
0.4	AC-AdHoc	Munich	AHAUC-029	Finalization of Phase 1 tests
0.5	SA4#34	Lisbon	S4-050188	Drafting of Phase 2 tests
0.6	SA4#35	San Diego	S4-050368	Finalization of Phase 2 tests

1 Introduction

This document contains the complete set of experimental designs for the subjective tests involved in the characterization phase of the 3GPP PSS/MMS(/MBMS) audio codec standardization. Based on the decision from SA#25, two codecs should be characterized, Enhanced aac-Plus (EAAC+) and Extended AMR-WB (AMR-WB+). The Characterization Test involves a number of specific tasks related to the subjective tests:

- Host Lab (HL) - processing the audio materials
- Mirror Lab (ML) - cross-checking of processed audio materials
- Listening Labs (LL) - conducting the listening tests and delivering raw data
- Global Analysis Lab (GAL) – collecting raw data, assembling the results, and drafting the Technical Report

The Characterization Test will be organized into a number of experiments using the MUSHRA [1] testing methodology. The amount of funds available for the Characterization Test (85.5k Euros) limits the number of experiments that may be conducted to eight MUSHRA tests. The experiments are subdivided into two phases of testing:

- Phase 1: Characterization of the two selected codecs across bit rates
- Phase 2: Characterization of the two selected codecs across error conditions

Phase 1 includes two experiments where each experiment is conducted by two listening labs (LL). Phase 2 includes four experiments where each experiment is conducted by one LL. The schedule for the two phases of testing for the audio codec characterization test is presented in Table 1.1.

Table 1.1 Schedule for conducting the Characterization Test (all dates in CY2005).

Schedule of tasks for the Phase 1 Experiments	
Mar.16	Codec proponents deliver executables to HL's
Mar.16	Selection of test items (subset of items used in the Selection Test)
Mar.16-Apr.4	Host Labs perform HL/Cross-check functions
Apr.4	HL's delivers processed materials to LL's
Apr.4-Apr.25	MUSHRA Listening tests (LL's)
Apr.25	LL's deliver raw voting data to GAL
Apr.25-May 6	GAL and draft TR preparation
May 9-May 13	Phase 1 results and draft TR presented at SA4#35
Schedule of tasks for the Phase 2 Experiments	
Jun. 7	Conference call to specify mapping of PDU into frames
Jun.14	Error patterns delivered to HL
Jun.14-Jul.12	Host Lab and mirror lab perform HL/Cross-check functions
Jul.12	HL's deliver processed materials to LL's
Jul.11-Aug.8	MUSHRA Listening tests (LL's)
Aug.8	LL's deliver raw voting data to GAL
Aug.8–Aug.27	GAL and final TR preparation
Aug.27-Sep.2	Review of results and TR
Sep.5-9	Phase 2 results and final TR presented at SA4#36

The experiments in Phase 1 will characterize the codecs across bit rates. The processing for all but one condition in the Phase 1 experiments will use the Floating-point Encoder and Fixed-point Decoder. The exception is the "AMR-WB+/low complexity" condition in experiment 1-1 for which the processing will use the Fixed-point Encoder and Fixed-point Decoder.

The experiments in Phase 2 will characterize the codecs across Packet Loss Rates (PLR).

- for Experiments 2-1 (Mono – EGPRS) and 2-2 (Stereo – EGPRS), Ericsson will provide error patterns on PDU level
- for Experiments 2-3 (Stereo – UTRAN – Lower bit rate) and 2-4 (Stereo – UTRAN – Higher bit rate), Qualcomm will provide error patterns

Tables 1.2 and 1.3 summarize the experiments and conditions involved in the Phase 1 and Phase 2 MUSHRA experiments, respectively.

Table 1.2. Experiments and conditions involved in Phase 1 the Characterization Test (Each experiment conducted in two listening labs).

Exp.1-1 - Mono	**bit-rate**
AMR-WB+	10k
aacPlus	10k
AMR-WB+	16k
aacPlus	16k
AMR-WB+	20k
aacPlus	20k
AMR-WB+ Low comp.	10k

Exp.1-2 Stereo	**bit-rate**
AMR-WB+	14k
aacPlus	14k
AMR-WB+	21k
aacPlus	21k
AMR-WB+	28k
aacPlus	28k

Table 1.3. Experiments and conditions involved in Phase 2 of the Characterization Test.

Exp. 2-1 EGPRS (Mono)	**bit-rate**	**PLR**
AMR-WB+	16k	0%
AMR-WB+	16k	1%
AMR-WB+	16k	6%
AMR-WB+	16k	10%
aacPlus	20k	0%
aacPlus	20k	1%
aacPlus	20k	6%
aacPlus	20k	10%

Exp. 2-2 EGPRS (Stereo)	**bit-rate**	**PLR**
AMR-WB+	24k	0%
AMR-WB+	24k	1%
AMR-WB+	24k	6%
AMR-WB+	24k	10%
aacPlus	24k	0%
aacPlus	24k	1%
aacPlus	24k	6%
aacPlus	24k	10%

Exp. 2-3 UTRAN (Stereo)	**bit-rate**	**PLR**
AMR-WB+	20k	0%
AMR-WB+	20k	1%
AMR-WB+	20k	5%
aacPlus	32k	0%
aacPlus	32k	1%
aacPlus	32k	5%

Exp. 2-4 UTRAN (Stereo)	**bit-rate**	**PLR**
AMR-WB+	40k	0%
AMR-WB+	40k	1%
AMR-WB+	40k	5%
aacPlus	40k	0%
aacPlus	40k	1%
aacPlus	40k	5%

Audio material to be used in the Characterization Test is classified according to the following three Audio Content types:

- Speech
- Music
- Mixed Music and Speech

In addition, the Mixed Content category is sub-classified into Speech over Music and Speech between Music. The distribution of categories for the test items is the same as that used in the Selection Test -- Speech - 4, Music - 4, Mixed - 4 (Speech over Music - 2, Speech between Music - 2).

The Characterization Test will use a subset of the Audio Materials that were used in the Selection test. Material selection will be performed by France Telecom R&D. The same set of test materials (4 training items and 12 grading items) will be used for all experiments in both Phases of the Characterization Test.

Table 1.4 provides an overview of the experiments and conditions included in the Characterization Test.

Table 1.4: Overview of Experiments in the Characterization Test

Phase	Exp.	Operational mode	# LL's	#Codecs	# cond/codec	#Anchors	#Ref.	# Signals	#items
1	1-1	Mono/bit-rate	2	2	3+1	2	2	11	12
1	1-2	Stereo/bit-rate	2	2	3	2	2	10	12
2	2-1	EGPRS/Mono/PLR	1	2	3	2	2	12	12
2	2-2	EGPRS/Stereo/PLR	1	2	3	2	2	12	12
2	2-3	UTRAN/Stereo/PLR	1	2	4	2	2	10	12
2	2-4	UTRAN/Stereo/PLR	1	2	4	2	2	10	12

Section 2 provides a list of test and reference codecs used in the experiments.

Section 3 provides a list of reference documents related to the test plan.

Section 4 specifies the kind of material to be used in the tests.

Section 5 gives general information relevant for all experiments.

Section 6 contains the test plan for the two phases of testing.

The specification of the processing functions of the audio material is included in section 7.

Annex A contains English language examples of instructions for the listening subjects for the MUSHRA tests to be carried out.

Annex B presents the filename convention.

1.1 Responsibilities

The funding for the characterization tests will use the funding already available (85.5 K Euro).

The processing and cross-check functions will be performed by Ericsson/Nokia and Coding Technologies without cost. Table 1.5 shows the responsibility for host lab functions for each experiment.

Table 1.5. Host Lab processing and cross-checking functions.

Exp.	Processing	Cross-check
1-1	Ericsson/Nokia	Coding Technology
1-2	Ericsson/Nokia	Coding Technology
2-1	Coding Technology	Ericsson/Nokia
2-2	Coding Technology	Ericsson/Nokia
2-3	Ericsson/Nokia	Coding Technology
2-4	Ericsson/Nokia	Coding Technology

The processing and cross-check laboratories will have the following responsibilities:

- Prepare testing and training material.
- Receive executables of the selected codecs from the codec proponents.
- Receive FER pattern files.
- Process reference, anchor and codec conditions (including re-sampling to sampling frequency of original material).
- Assemble the final distribution of the processed material to the listening laboratories.

The characterization experiments will be run by the LL's as shown in Table 1.5. The compensation for each listening test will be the same as that for the Selection Test – 9k Euros.

Table 1.5: Allocation of sub-experiments to the Listening Laboratories

Exp.	Listening Labs	
1-1	Dynastat	Ericsson
1-2	FT-R&D	Cod.Tech.
2-1	Fraunhofer	---
2-2	NTT-AT	---
2-3	Nokia	---
2-4	T-Systems	---

Each LL will be required to provide a full report of the experiments performed. The test results and raw voting data will be delivered to the Global Analysis Laboratory (GAL) in spreadsheets prepared by the GAL for that purpose. Any deviations from the listening test specifications contained in this document will be documented by the LL along with the results.

The test results will be combined by the GAL (Dynastat) and presented to SA4. The GAL will also draft the Technical Report for the remaining funding of 13.5k Euros. Specifically, the GAL will perform the following tasks:

- Prepare and deliver the randomizations of training and test items for each LL
- Prepare and provide spreadsheets to each LL for delivery of raw MUSHRA grading data
- Assemble and combine results from LL's for each experiment
- Perform appropriate statistical analyses to:
    - Compare results between LL's
    - Provide summary results for each experiment
- Prepare and present the Technical Report

1.2 Test codecs

Table 1.6 provides an overview of the selected test codecs participating in the PSS/MMS/MBMS audio codec characterization test.

Table 1.6. Codecs involved in the Characterization Test

Codec	Providing Organization(s)	Contact(s)
Enhanced aacPlus	Coding Technologies	kunz@CODINGTECHNOLOGIES.COM
Extended AMR-WB+	Ericsson/Nokia/VoiceAge	Stefan.Bruhn@ericsson.com pasi.s.ojala@nokia.com RedwanS@VOICEAGE.COM

1.3 Anchors and references

In addition to the items encoded with the test codecs, anchor and reference items will be included in the tests. Their purpose is to normalize the tests and to make them more comparable across different experiments and in different LL's.

Both the mono and the stereo experiments will include two anchors (low-pass filtered original signal at 3.5kHz and 7.0kHz) as specified in the MUSHRA testing methodology. The unprocessed original source will also be included in each experiment, once as the “open reference” and once as the “hidden reference” as specified in the MUSHRA testing methodology. Table 1.7 shows the standard reference conditions used in the experiments.

Table 1.7. Standard reference conditions involved in the MUSHRA tests.

#	Type	Specification	Channel type
1	Open Reference	Original source signal	Mono/Stereo as appropriate
2	Hidden Reference	Original source signal	Mono/Stereo as appropriate
3	Anchor	3.5 kHz Lowpass Original signal	Mono/Stereo as appropriate
4	Anchor	7.0 kHz Lowpass Original signal	Mono/Stereo as appropriate

2 References and Conventions

2.1 Reference Documents

[I] RECOMMENDATION ITU-R BS.1534 Method for the subjective assessment of intermediate quality level of coding systems

3 Test Material

The test material will be composed of a subset of the material selected for use in the selection test.

3.1 Training material

Limited material will be used in the training phase in which the subjects familiarize with the testing methodology and environment.
The training will include four items, one from each audio signal category. These items will be identified by the selection entity and shall not be re-used in the grading phase. The training phase will be executed as a separate short MUSHRA session.

4 Information relevant to all Experiments

4.1 General Technical Notes

Any and all deviations from the specifications contained in this document must be documented and submitted to TSG-SA-WG4 along with the experimental results.

For all experiments, subjects should be seated in a quiet environment; 30dBA Hoth Spectrum (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) measured at the head position of the subject. This will help ensure consistency between the different subjects in the same laboratory as well as across the different laboratories in which these experiments will be performed.

The test stimuli will be presented to the subjects over headphones meeting one of the following requirements:

1) Binaural listening using closed-back, supra-aural headphones.
2) Binaural listening using open-back, circum-aural headphones.

4.2 Testing methodology

The testing is carried out according to MUSHRA methodology [1], which is suitable for evaluation of intermediate audio quality and gives accurate and reliable results. The labs carrying out the testing should have experience with the MUSHRA method from earlier exercises. The MUSHRA test method applied here uses the original unprocessed material with full bandwidth as the reference signal (which is also used as a hidden reference), two hidden anchors, the conditions of the codecs under test, and the reference conditions with which the codecs under test are to be compared.

4.3 Error Patterns and Error Conditions

Error conditions will be applied in experimental block B according to AMR-WB+ performance requirements, v. 2.0 (Tdoc S4-030434) and PSS/MMS Audio Codec Selection, Design Constraints and Performance Requirements – Version 2.0. Details on the availability of the error patterns and their application are given in section 8 (Processing). .

4.4 Training phase

Prior to the actual testing, a training phase is carried out in which the test subjects are familiarized with the testing methodology and the testing environment. The training is conducted using the same MUSHRA methodology as the actual test, though training is limited to four trials.

The training phase is based on the same codec, anchor, and reference conditions as the grading phase.

4.5 Selection of subjects

The selection of subjects follows the guidelines given in [1]. In particular, it is recommended that experienced listeners should be used. These listeners should have some experience in listening to sound in a critical way. Such listeners will give more reliable results and in a more timely manner than non-experienced listeners.

4.5.1 Screening of subjects

Sometimes there is justification for using a subject rejection procedure either before (pre-screening) or after (post-screening) the test data has been collected. In some cases both types of rejection procedures can be used. In these cases, subject rejection is a process where all judgments from a particular subject are omitted from the results of the experiment.
The employment of rejection procedure can lead to biased results. Therefore, it is important to both justify and describe subject rejection procedures in the LL test reports.

4.5.1.1 Pre-screening of subjects

The listening panel should be composed of experienced listeners, in other words, people who understand and have been properly trained in the described method of subjective quality evaluation. These listeners should:

- have experience in listening to sound in a critical way;
- have normal hearing (ISO Standard 389 should be used as a guideline).

The training procedure might be used as a tool for pre-screening.

4.5.1.2 Post-screening of subjects

Post-screening methods can be roughly separated into at least two classes:

- one is based on the ability of the subject to make consistent repeated votes;
- the other relies on inconsistencies of an individual grading compared with the mean result of all subjects for a given item.

It is recommended to look to the individual spread and to the deviation from the mean grading of all subjects. The aim of this is to get a fair assessment of the quality of the test items. If few subjects use either extreme end of the scale (excellent, bad) and the majority are concentrated at another point on the scale, these subjects could be recognized as outliers and might be rejected.

Due to the fact that “intermediate quality” is tested, a subject should be able to identify the coded version very easily and therefore find a grade which is in the range of the majority of the subjects. Subjects with grades at the upper end of the scale are likely to be less critical and subjects who have grades only at the lowest end of the scale are likely to be too critical. By rejecting these extreme subjects a more realistic quality assessment is expected.
The methods are primarily used to eliminate subjects who cannot make the appropriate discriminations. The application of a post-screening method may clarify the tendencies in a test result. However, bearing in mind the variability of subjects’ sensitivities to different artifacts, caution should be exercised.

Taking into account the size of the listening panel used throughout the experiments, the effects of any individual subject’s grades is low and so the need to reject a subject’s data is greatly diminished.

5 Phase 1 Experiments

5.1 Introduction

The experiments in this block are designed to evaluate the error-free, generic audio signal performance of the selected codecs under ideal conditions.

The experimental block covers two experiments:
- Mono – 3 bit rates for 2 codecs + 1 low complexity condition for AMR-WB+
- Stereo - 3 bit rates for 2 codecs

The details provided in this section are those that are specific to this particular experiment. Generic information, relevant to this and other experiments can be found in Section 4. Therefore Listening Laboratories should use the information in Section 4 in conjunction with the information provided in this section.

5.2 Test Conditions

Tables 5.1 and 5.2 provide an overview of the conditions applicable to the Phase 1 experiments.

Table 5.1: Conditions and factors for Experiment 1-1 (mono – bit rate)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Mono
Bit rates	3	10kbps, 16kbps, 20kbps
Low complexity-AMR-WB	1	10kbps
References		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	3.5 kHz and 7 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		48 kHz
Number of input channels	1	Mono input
Output sampling rate		Unspecified
Number of output channels	1	Mono
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Replications	2	Each sub-experiment is done in two independent test labs
Listening System	1	Binaural high-quality headphones
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)

Table 5.2: Conditions and factors for Experiment 1-2 (stereo - bit-rate)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Stereo
Bit rates	3	14kbps, 21kbps, 28kbps
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	7 kHz low-pass filtered original signal 3.5 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		48 kHz
Number of input channels	2	Stereo input
Output sampling rate		Unspecified
Number of output channels	2	Stereo
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Replications	2	Each sub-experiment is done in two independent test labs
Listening System	1	Binaural high-quality headphones
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)

5.3 Material

See section 3.

5.4 Experimental Design

See section 4.2

5.5 Opinion Scale

The question asked of the subject will be a continuous Listening Quality Scale ranging from 0 to 100. The intervals 0 to 20 correspond to BAD, 20 to 40 to POOR, 40-60 to FAIR, 60 to 80 to GOOD, and 80 to 100 to EXCELLENT.

5.6 Processing

Processing is specified in section 7.

5.7 Duration of the Experiment

The duration of the experiment per subject depends on the number of trials and on the number of items per trial. However, it can be assumed that each vote requires listening to the respective item, the open reference and two (quality-wise) neighboring items 2 times. With an assumed average length per item of 7.5 s, the test will consume a listening time per subject of:
#trial * #hidden items/trial * (1+1+2) * #re-listenings * length/item.

For the grading phase in each of the sub-experiments the estimated duration is" 12 * 8 * 4 * 2 * 7.5s = 1.6 hours per subject

For the training phase the number of re-listenings is assumed to be 1. For each of the sub-experiments this accounts to
4 * 8 * 4 * 1 * 7.5s = 16 min per subject

In order to avoid listener fatigue, sufficient breaks are required between the trials.

The experiments can be carried out with several subjects in parallel provided that a corresponding number of proper listening facilities are available.

5.8 Votes Per Condition

The number of votes per conditions is identical with the number of subjects per sub-experiment.

5.9 Randomizations

Each listener will be presented with the sound items in an individual random presentation order. Also the order of the trials will be random per individual.

6 Phase 2 Experiments

6.1 Introduction

The experiments in this block are designed to evaluate the audio signal performance of the selected codecs under stressed operating conditions.

The experimental block covers four experiments:

- EGPRS Error conditions (mono)
- EGPRS Error conditions (stereo)
- UTRAN Error conditions (stereo-low bit-rate)
- UTRAN Error conditions (stereo-high bit-rate)

The details provided in this section are those that are specific to this particular experiment. Generic information, relevant to this and other experiments can be found in Section 4. Therefore Listening Laboratories should use the information in Section 4 in conjunction with the information given in this section.

6.2 Test Conditions

Tables 6.1 through 6.4 provide an overview of the conditions applicable to the Phase 2 experiments.

Table 6.1: Conditions and factors for Experiment 2-1 (EGRPS, Mono, Four Packet Loss Rates)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	B (MMS)
Error contirions	4	0%, 1%, 6%, and 10% Packet Loss Rates
Mono/Stereo	1	Mono
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	3.5 kHz and 7 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		16 kHz
Number of input channels	1	Mono input
Output sampling rate		16 kHz
Number of output channels	1	Mono
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Replications	1	Each sub-experiment is done in one independent test lab
Listening System	1	Binaural high-quality headphones
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)

Table 6.2 Conditions and factors for Experiment 2-2 (EGPRS, stereo, four Packet Loss Rates)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	B (MMS)
Error Conditions	4	0%, 1%, 6%, and 10% Packet Loss Rates
Mono/Stereo	1	Mono
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	7 kHz low-pass filtered original signal, 3.5 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		48 kHz
Number of input channels	2	Stereo input
Output sampling rate		Unspecified
Number of output channels	2	Stereo
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
Replications	1	Each sub-experiment is done in one independent test lab
Listening System	1	Binaural high-quality headphones
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)

Table 6.3 Conditions and factors for Experiment 2-3 (UTRAN, stereo, three Packet Loss Rates)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	B (MMS)
Error Conditions	4	0%, 1%, and 5% Packet Loss Rates
Mono/Stereo	1	Mono
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	7 kHz low-pass filtered original signal, 3.5 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		48 kHz
Number of input channels	2	Stereo input
Output sampling rate		Unspecified
Number of output channels	2	Stereo
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
Replications	1	Each sub-experiment is done in one independent test lab
Listening System	1	Binaural high-quality headphones
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)

Table 6.3 Conditions and factors for Experiment 2-4 (UTRAN, stereo, three Packet Loss Rates)

Main Codec Conditions		
Codec(s)	2	AMR-WB+ and EAAC+
Use case	1	B (MMS)
Error Conditions	3	0%, 1%, and 5% Packet Loss Rates
Mono/Stereo	1	Stereo
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	2	7 kHz low-pass filtered original signal, 3.5 kHz low-pass filtered original signal
Common Conditions		
Stimulus type		Sound item
Radio Channels	0	Clean
Number of audio items	12	
Input sampling rate		48 kHz
Number of input channels	2	Stereo input
Output sampling rate		Unspecified
Number of output channels	2	Stereo
Listening Level	1	To be chosen by subject
Listeners	15	Experienced listeners
Presentation randomizations	15	One for each listener
Rating Scale	1	Continuous quality scale
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
Replications	1	Each sub-experiment is done in one independent test lab
Listening System	1	Binaural high-quality headphones
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)

6.3 Material

See section 3.

6.4 Experimental Design

See section 4.2

6.5 Opinion Scale

The question asked of the subject will be a continuous Listening Quality Scale ranging from 0 to 100. The intervals 0 to 20 correspond to BAD, 20 to 40 to POOR, 40-60 to FAIR, 60 to 80 to GOOD, and 80 to 100 to EXCELLENT.

6.6 Processing

Processing is specified in section 7.

6.7 Duration of the Experiment

The duration of the experiment per subject depends on the number of trials and on the number of items per trial. However, it can be assumed that each vote requires listening to the respective item, the open reference and two (quality-wise) neighboring items 2 times. With an assumed average length per item of 7.5 s, the test will consume a listening time per subject of:
#trial * #hidden items/trial * (1+1+2) * #re-listenings * length/item.

For the grading phase in each of the sub-experiments this accounts to
12 * 8 * 4 * 2 * 7.5s = 1.6 hours per subject

For the training phase the number of re-listenings is assumed to be 1. For each of the sub-experiments this accounts to
4 * 8 * 4 * 1 * 7.5s = 16 min per subject

In order to avoid listener fatigue, sufficient breaks are required between the trials.

The experiments can be carried out with several subjects in parallel provided that a corresponding number of proper listening facilities are available.

6.8 Votes Per Condition

The number of votes per conditions is identical with the number of subjects per sub-experiment.

6.9 Randomizations

Each listener will be presented with the sound items in an individual random presentation order. Also the order of the trials will be random per individual.

7 Processing

Common for the processing of all conditions is that it will be done with concatenated material. To that purpose in a pre-processing step, for each experiment the material required for it including training material will be concatenated to one single file. This file will then in the main-processing step(s) be processed through the respective candidate or reference codecs, or the specific anchor signal processing will be applied. After processing and proper compensation of the processing delay, the material will finally be sampling-rate converted and split-up again to separate items.

7.1 Pre-processing

In general, the material selection panel is responsible for selection of proper training and test material. However, for clean speech items, a particular pre-processing is necessary for removing possible silence segments included in the original speech item files.
After this speech-specific pre-processing the complete material will be concatenated to one single file.

7.1.1 Clean speech items

The pre-processing for the speech items has to make sure that leading and trailing silence segments are removed. A corresponding program tool will be provided by:
Volunteering organization Ericsson

The host lab will do the pre-processing applying that script.

7.1.2 Concatenation of material

The concatenation will be done using an executable script. The script will take as input all items of the training material and the test material belonging to the respective experiment. The order in the concatenation is such that the training material will precede the test material. In order to ensure that the last items in the concatenated file will not be clipped as cause of frame-wise processing or codec delay, a short artificial silence item will be added to the end of the concatenated file. Output of the concatenation script will be a single file containing the concatenated material and a time-file comprising name identifiers and time markers of the individual items. The time-file will be used after processing for split-up of the material to the individual items.

The concatenation script will be provided by:
Volunteering organization Ericsson.

7.2 Main processing

7.2.1 Processing of anchor conditions, 16 kHz downsampling, stereo to mono mixing

The anchor conditions as well as re-sampling and stereo to mono mixing are performed using a script making use of tools of the AFSP library such as “ResampAudio” and “CopyAudio”.

The tools are freely available on the Internet, the processing script and a copy of the required tools (including usage assistance, if necessary) will be provided by Coding Technologies.
The command line syntax of the processing script is as follows:

anchor –lp<cut-off> –s<stereo_degree> –fsout16000 –monoout <infile.wav> <outfile>,
where –lp<cut-off>, –s<stereo_degree>, –fsout16000, and –monoout are optional arguments and

<cut-off> = {3500, 7000}	is the cut-off frequency of the low-pass to be used, and
<stereo_degree> = {6, 12}	is the side channel attenuation in dB.

The argument –fsout16000 is supplied if the signal has to be re-sampled to 16 kHz sampling frequency.
The argument –monoout is supplied if the output signal has to be a one-channel mono-signal.

The input file is assumed to be of wav format.

The output file format is wav if the name has the extension .wav. If it contains the extension .raw, then a headerless 1-channel (mono) 16 bit PCM file is generated.

7.2.2 Codec(s) and reference codec processing

The concatenated material file will be created by the host lab. After processing, the encoded material as well as the concatenated original files will be delivered to codec proponent(s) for cross-checking. The proponent(s) make sure that possible codec delay is properly compensated. The procedure of how the delay was compensated need to be reported as part of the respective codec description(s).

7.2.2.1 Impaired channel processing

The host lab does the impaired channel processing with a FER pattern file identifying correct and erroneous frames. The host lab will produce that file using a script provided by:
Volunteering organization: <tbd>

The input to the script is a seed to the random generator included in the script. It will be provided by ETSI Secretary, Paolo Usai after candidate submission.
The format of the FER pattern file is ASCII text with one line per frame. A ‘0’ (zero) in a line indicates a correct frame, a ‘1’ (one) indicates a frame erasure.
The length of the file will be sufficient to cover the complete concatenated training and test material, assuming a minimum frame length of 10 ms. Note, that depending on the actual frame size of the individual candidate codecs, the FER pattern file will not be applied in its entire length.

In order to realize the randomization of the error patterns across each individual listening (see paragraph 7.9), before application the error pattern file will be circularly shifted with individual time offsets. The time offset (in frames) for listener *s*, lab *p* and experiment *e* (B3=1, B4=2) is calculated as follows:
$offs = \mathrm{mod}(s \cdot 12345 + p \cdot 31415 + e \cdot 27183, totlen)$, where *totlen* is the total length of the FER pattern file (in frames).
The circular shift will be done using an executable script with the command line syntax:
shiftcirc –<offs> <inpat> <outpat>
The script will be provided by Ericsson.

7.3 Post-processing

The post-processing comprises the steps of up-sampling to the original sampling frequency of 48 kHz and split-up of the concatenated material.

7.3.1 Up-sampling

Resampling to the original sampling frequency of 48 kHz will be done using a script with the following command-line syntax:

upsamp48 –fs<fsamp> <infile> <outfile.wav>

The script will make use of AFSP tools such as “ResampAudio”.

In order to avoid possible signal clipping distortions introduced after re-sampling during the conversion of the internal floating-point to the 16-bit integer representation of the wav-file, before this conversion the re-sampled signal is attenuated by applying a constant factor of 0.93 (≈ -0.3 dB).
Note, in order to ensure that this factor is consistently applied to all codec, anchor and reference conditions, the program must always be applied, even if actually no re-sampling is necessary since the sampling rate of the respective signal is already 48 kHz.

The optional argument –fs<fsamp> to the processing script needs to be supplied in case of a PCM input signal and specifies the sampling frequency in Hz.

The input file format is wav if the name has the extension .wav. If it has the extension .raw, then a headerless 1-channel (mono) 16 bit PCM file is assumed. In this latter case, the –fs option has to be supplied.

The tools are freely available on the Internet, the processing script and a copy of the required tools (including usage assistance, if necessary) will be provided by Coding Technologies.

7.3.2 Split-up of processed material

The split-up of the processed material to partial sound files will be done using a script which is functionally inverse to the concatenation procedure described in section 8.1.2. The program will take the concatenated and processed (and delay-compensated) file along with the time-file as input. On time intervals of length 2*N* around the time markers n_i indicated by the time-file it will first apply a roll-off/roll-on window with Hanning characteristic:

$$x_w(n) = x(n) \cdot \sin^2\left(\frac{\pi}{2N}(n - n_i)\right)$$

and then split-up the file at the time markers n_i. *N* corresponds to a roll-off/roll-on time of 100 ms. I.e. N=0.1 * f_s where f_s = 48 kHz. Partial sound files will be generated with names according to the name identifiers comprised in the time-file and a processing tag, identifying the kind of processing.
This actual split-up and windowing is performed using the *astrip* program from the ITU-T STL (STL 2000):

Source:	ITU-T
Location:	STL sub-directory: unsupported
Format:	C source code
Program:	*astrip*

The split-up script will have the following command-line syntax:

splitup <infile> <processing_tag> <timefile>

The processing tag is an arbitrary string which is added to the name identifiers given in the time-file.

The generated output files are of wav format.

The script will be provided by Ericsson

Annex A: Listeners Instructions

The following is an example of the type of instructions that should be given to or read to the subjects in order to instruct them on how to perform the test.

Instructions to be given to subjects

1 Training phase

The first step in the listening tests is to become familiar with the testing process. This phase is called a training phase and it precedes the formal evaluation or grading phase.

The purpose of the training phase is to allow you, as an evaluator, to learn how to use the test equipment and the grading scale.

In the training phase you will participate a short test similar to the one you will perform in the grading phase of the test.

No grades given during the training phase will be taken into account in the actual tests.

2 Grading phase

The purpose of the grading phase is to invite you to assign your grades using the quality scale. Your grades should reflect your subjective judgment of the quality level for each of the sound excerpts presented to you. Each trial will contain <x> signals to be graded. Each of the items is approximately 5 to 10 s long. You should listen to the reference and all the test conditions by clicking on the respective buttons. In a first step it is recommended to browse through all signals within each trial in order to get a coarse impression of the offered quality range. Then you may listen more carefully and start to rank the signals. You may listen to the signals in any order, any number of times. Use the slider for each signal to indicate your opinion of its quality. When you are satisfied with your grading of all signals you should click on the “register scores” button at the bottom of the screen.

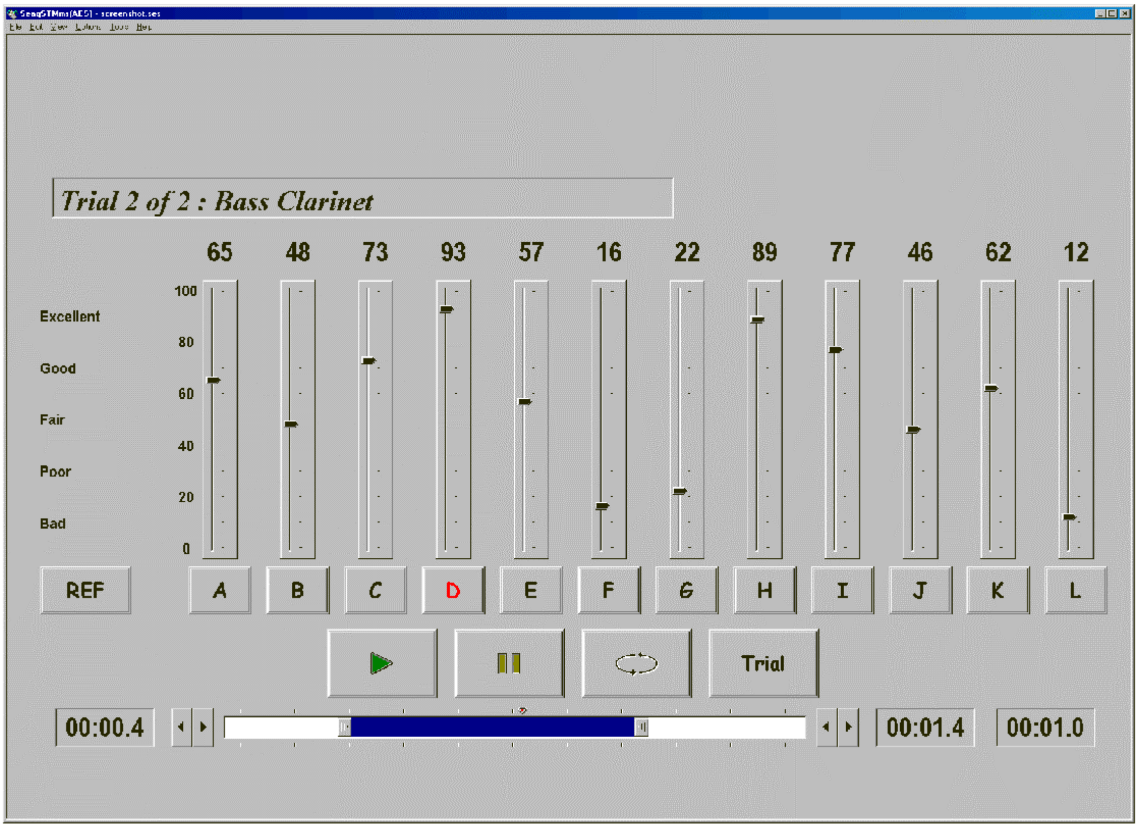
You will use the quality scale as given in Fig. 1 when assigning your grades.

The grading scale is continuous from “excellent” to “bad”. A grade of 0 corresponds to the bottom of the “bad” category, while a grade of 100 corresponds to the top of the “excellent” category.

In evaluating the sound excerpts, please note that you should not necessarily give a grade in the “bad” category to the sound excerpt with the lowest quality in the test. However one or more excerpts must be given a grade of 100 because the unprocessed reference signal is included as one of the excerpts to be graded.

You should not discuss your personal interpretation or your gradings with the other subjects at any time during the test.

An example of the user interface used in the blind grading phase



1534-04

Annex B: File formats, Naming conventions, Directory structures, Platforms

B-1 File formats

Unless specified differently all signal files are of wav format using a linear 16-bit PCM sample representation. The sampling rate prior to pre-processing and after post-processing is 48 kHz.

B-2 Computer platform

The processing will be done in a Cygwin environment under Windows.

B-3 Directory structure

The suggested directory structure to be used for processing is given below. It is though the freedom of the host labs to choose other, possibly more appropriate directory structures.

```
Root
|
-.- org                    Original (unprocessed) material
 .- preproc                Pre-processed material
 .- ep                     Error pattern and program to generate it
 .- bin                    Programs  and  scripts  required  for  pre-processing
                           and error pattern generation
 |
 .- cand_1                 Directory for candidate codec 1
 |   |
 |   .- processed          processed material
 |   .- bin                programs and scripts required for processing
 |   .- tmp                intermediate processing results
 |
 .- cand_2                 Directory for candidate codec 2
 |   |
 |   .- processed          processed material
 |   .- bin                programs and scripts required for processing
 |   .- tmp                intermediate processing results
 |
 .- cand_3                 Directory for candidate codec 3
 |   |
 |   .- processed          processed material
 |   .- bin                programs and scripts required for processing
 |   .- tmp                intermediate processing results
 |
 .- ref                    Reference codecs, anchors and hidden references
 |   |
 |   .- processed          processed material
 |   .- bin                programs and scripts required for processing
 |   .- tmp                intermediate processing results
 |
 .- output                 Processed material per sub-experiment
     |
     .- exp_A1a
     .- exp_A1b
     .- exp_A2a
     .- exp_A2b
     .- exp_A3a
     .- exp_A3b
     .- exp_A4a
     .- exp_A4b
     .- exp_B1a
```

```
.- exp_B1b
.- exp_B2a
.- exp_B2b
.- exp_B3a
.- exp_B3b
.- exp_B4a
.- exp_B4b
```

.

B-4 File naming

A suggested file naming convention is given below. It is though the freedom of the host labs to choose other, possibly more appropriate naming conventions as long as they are well documented and unambiguous.

1. Unprocessed original signals:

<main_category>_<sub_category>_<stereo_category>_<item_no>_org.wav

where

- o <main_category> is one out of {m, som, sbm, s} (for music, speech over music, speech between music, speech),
- o <sub_category> is one out of
  - o {si, vo, ch, po, cl, ot} for music main category (single instrument, vocal, choir, pop, classical, other)
  - o {ne, ad, fi, ot} for speech over music main category (news trailer, advertisement, film sound track, other)
  - o {sm, ms, js, sj, as, ot} for speech between music main category, and
  - o {cl, no} for speech main category (clean, noisy)
- o <stereo_category> is one out of {2t, mt, ft} for the speech category tested in the stereo sub-experiments (two talkers, moving talker, fixed talker) and {x} for all other items without explicit stereo category.
- o <item_no> is a number identifying the item out of the respective main and sub-category.

2. Pre-processed original signals after (possible) speech-silence clipping:

<main_category>_<sub_category>_<stereo_category>_<item_no>.wav

Note, that even for non-speech files for which no silence-clipping will be applied files according to this naming have to be created (by copy operation) as these files will serve as input files to the sub-sequent concatenation step.

3. Pre-processed original signals after concatenation:

all_cat.wav

The corresponding time-file comprising the segmentation information is:

all_cat.tim

4. Material after main processing:

all_cat_<codec_id>_<exp_id>.wav

where

- o <codec_id> is one out of

    - {cand_1, cand_2, cand_3, AAC, AMRWB, lp3500_s12, lp7000_s12, lp7000_s6, hidref, opref} for the experiments involving stereo and
    - {cand_1, cand_2, cand_3, AAC, AMRWB, lp3500, lp7000, hidref, opref} for the experiments in mono.
    -
- <exp_id> is one out of {A1, A2, A3, A4, B1, B2, B3, B4}

5. Processed material after de-concatenation:

<main_category>_<sub_category>_<stereo_category>_<item_no>_<codec_id>_<exp_id>.wav

6. Processed material after allocation to sub-experiments:

<main_category>_<sub_category>_<stereo_category>_<item_no>_<codec_id>_<exp_id>_<sub_exp>.wav

where <sub_exp> is one out of {a, b}.

7. Processed material

<main_category>_<sub_category>_<stereo_category>_<item_no>_<exp_id>_<sub_exp>_<cond_id>.wav

where <cond_id> is one out of

- {cond[1-8], opref} for the mono experiments and
- {cond[1-10], opref} for the stereo experiments.