Technical Specification Group Services and System Aspects
Meeting #21, Frankfurt, Germany, 22-25 September 2003

TSGS#21(03)0437

3GPP TSG-SA4 meeting
September 1-5, 2003, Erlangen, Germany

Tdoc S4-030700

Title: **AMR-WB+ and PSS/MMS Low-Rate Audio Selection Test and Processing Plan Version 2.0**

Source: **TSG SA WG4**

Document for: **Information**

Agenda Item: **7.4.1**

Contents

0 Revision History

V 0.2: Added new test plans for experiments 2-4.
V 0.3: Major change after agreement on joint PSS/MMS low-rate audio and AMR-WB+ selection testing
V 0.4: Changes during review in joint low-rate audio ad-hoc session.
Reviewed sections: 1-4, 8
V 0.5: Updates in section 5.
Sections 6 and 7 aligned with agreed tables 1-1 and 1-2, 20 instead of 10 listeners.
Section 7: Randomization of error patterns across all listenings
Updates in processing section 8.
Annex A: Listener instructions added
V 0.5.1: Remaining issues:
16 kHz input sampling freq
material selection
stereo anchors
directory structure and naming conventions
solved issues:
training phase
V 0.5.2: Remaining issues:
16 kHz input sampling freq
stereo anchors
directory structure and naming conventions
solved issues:
material selection
V 0.6: Remaining issues:
16 kHz input sampling freq
stereo anchors tbc.
directory structure and naming conventions

V 0.7: Stable draft
V 0.7R: cleaned version, minor editorial changes

Remaining issues:

Editorial updates

Verification of processing and naming conventions

Identify entities:

- o Listening labs
- o Host lab and mirror host lab
- o Selection entity
- o Entity providing vote collection template
- o Global analysis lab
- o Volunteers providing processing scripts and tools

Actions required to identify the entities:

- o Draft letter calling for entities (NEC)
- o Listening labs:
  SA4 endorsed call for potential listening labs
- o Host lab and mirror host lab
  SA4 endorsed call for potential labs
- o Selection entity
  SA4 endorsed call for potential entities
- o Entity providing vote collection template
  SA4 endorsed call for potential entity
- o Global analysis lab
  SA4 endorsed call for potential entity
- o Volunteers providing processing scripts and tools
  Assignment of entities by correspondence or still during SA4#27

V 0.8:
Allocation of listening and host labs
Updates to Processing section
Remaining issue:
Identify entity providing scrpts for impaired channel processing

V 0.9:
Allocation of listening and host labs
Updates to Processing section

Remaining issues:
Identify entity providing scripts for speech pre-processing
Identify entity providing scripts for impaired channel processing
Identify Material selection entity
List contact names to all involved entities

1 Introduction

This document contains the complete set of experimental designs for the selection phase of the Wideband Adaptive Multi-Rate extension codec (AMR-WB+) as well as of the PSS/MMS low-rate audio codec selection. The experiments are designed such that they are relevant for both exercises.

The experiments are subdivided into two main blocks:
Experiments A: Intrinsic quality comparison of candidate codecs
Experiments B: Quality comparison under stressed operating conditions

Experimental block A will test the codecs in the following operational modes:

- 14 kbps, mono, use case A (PSS)
- 18 kbps, stereo, use case A (PSS)
- 24 kbps, mono, use case A (PSS)
- 24 kbps, stereo, use case A (PSS)

Experimental block B will test the codecs under the following operational conditions:

- 14 kbps, mono, use case B (MMS), 16 kHz input and output sampling rate.
- 18 kbps, stereo, use case B (MMS),
- 14 kbps, mono, use case A (PSS), 3% FER
- 24 kbps, stereo, use case A (PSS), 3% FER

Operational modes/conditions not covered by experimental blocks A and B, for which performance requirements are defined, will be tested during characterization of the selected codec.

Audio material classified according to the following four content types will be used in all of the experiments:

- Speech
- Speech over Music
- Speech and Music interlaced
- Music

The experiments for each operational mode/condition further divide into 2 sub-experiments of manageable size with a mix of material out of each content type class.

The following tables provide an overview of experimental block A and B:

Table 1-1: Sub-experiments of experimental block A

Exp.	Operational mode	Audio Material	#Codecs in test	# reference codecs	#Anchors in test	#Referen ces	#items	Total
A1a	14 kbps, mono, use case A (PSS)	Set a	3	2	2	2	12	108
A1b		Set b						
A2a	18 kbps, stereo, use case A (PSS)	Set a	3	2	3	2	12	120
A2b		Set b						
A3a	24 kbps, mono, use case A (PSS)	Set a	3	2	2	2	12	108
A3b		Set b						
A4a	24 kbps, stereo, use case A (PSS)	Set a	3	2	3	2	12	120
A4b		Set b						

Table 1-2: Sub-experiments of experimental block B

Exp.	Operational condition	Audio Material	#Codecs under test	# reference codecs	#Anchors in test	#References	#items	Total
B1a	14 kbps, mono, use case B (MMS), 16 kHz input and output sampling rate	Set a	3	2	2	2	12	108
B1b		Set b						
B2a	18 kbps, stereo, use case B (MMS)	Set a	3	2	3	2	12	120
B2b		Set b						
B3a	14 kbps, mono, use case A (PSS), 3% FER	Set a	3	2	2	2	12	108
B3b		Set b						
B4a	24 kbps, stereo, use case A (PSS), 3% FER	Set a	3	2	3	2	12	120
B4b		Set b						

Section 2 provides a list of candidate 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 along with a material selection procedure.

Section 5 gives general information relevant for all experiments.

Sections 6 and 7 contain the test plans for the two experimental blocks.

The specification of the processing functions of the audio material is included in section 8

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

Annex B presents the filename convention.

1.1 Responsibilities

The processing will be done by the host-lab T-Systems and Audio Research Labs (mirror lab) and verified by the candidates. The funding for the Selection and the characterization tests will be equally shared between the proponents.

T-Systems will serve as host laboratory. The host laboratory will have the following responsibilities:

- Receive candidate testing material after the call for the material and distribute it to the material selection panel.
- Prepare testing and training material based on the indications from the material selection panel.
- Receive executables of the candidate codecs from the codec proponents.
- Receive executables of the reference codecs.
- Generate FER pattern file after receipt of the random generator seed.
- Process reference, anchor and codec conditions (including re-sampling to sampling frequency of original material).
- Cross-check the processing between the host laboratories and the codec proponents: provide the processed material to the codec proponents who cross-check.
- Assemble the final distribution of the processed material to the listening laboratories, which includes blinding of the material.

Listening labs:
Fraunhofer Geselschaft (FhG), France Telecom (FT) , T-Systems (TS), NTT-AT, Dynastat (D), Nokia (N), Ericsson (E), Coding Technologies (CT).

The selection experiments for each candidate codec are run by two listening laboratories in parallel, as shown in Table 1-2.

Table 1-2: Allocation of sub-experiments to the Listening Laboratories

Exp.	Lab1	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7	Lab8	Total
LL ID	FhG	CT	E	N	D	FT	TS	NTT_AT	Per Exp.
A1a	x				X				2
A1b		x				X			2
A2a			x				x		2
A2b				x				x	2
A3a	x				X				2
A3b		x				X			2
A4a			x				x		2
A4b				x				x	2
B1a	x				X				2
B1b		x				X			2
B2a			x				x		2
B2b				x				x	2
B3a	x				X				2
B3b		x				X			2
B4a			x				x		2
B4b				x				x	2
Totals:	4	4	4	4	4	4	4	4	32

Each listening lab will be asked to provide a full report of the experiments performed. The test results will be included in spreadsheets prepared to that purpose. Any deviations from the specifications contained in this document will be documented along with the results.

The test results will then be combined by the Global Analysis Laboratory (Audio Research Labs) and presented to SA4.

1.2 Contact Names

Tbd.

2 Codecs, Anchors and References

2.1 Candidate and Reference codecs

The experiments are designed to be able to accommodate 4 candidate codecs. The number may however simply be adjusted downwards, if, e.g. candidates decide to withdraw or decide not to be part in certain parts of the experiments.
The following table provides an overview of the candidate codec participating in the AMR-WB+ and PSS/MMS low-rate audio selection tests.

#	Codec name	AMR-WB+ candidate	PSS/MMS low-rate audio candidate	Providing Organization(s)	Contact
1	AAC+	No	Yes	Coding Technologies/ NEC	kunz@CODINGTECHNOLOGIES.COM Frederic.Gabin@NECTECH.FR
2	AMR-WB+	Yes	Yes	Ericsson/ Nokia/ VoiceAge	Stefan.Bruhn@ericsson.com pasi.s.ojala@nokia.com VesaR@VOICEAGE.COM
3	CT	No	Yes	Coding Technologies	kunz@CODINGTECHNOLOGIES.COM

The reference codecs are listed in the following table.

#	Codec name	AMR-WB+ candidate	PSS/MMS low-rate audio candidate	Providing Organization(s)	Contact
5	AAC (Fraunhofer)	No	No	Fraunhofer	Bernhard.Grill@iis.fraunhofer.de Johannes.Hilpert@iis.fraunhofer.de
6	AMR-WB	No	No	3GPP	pasi.s.ojala@nokia.com

2.2 Anchors and references

Besides the items encoded with the candidate and reference 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 realizations.
In the experiments testing mono signals two anchors will be used with lowpass filtered original signal.
In the experiments testing stereo signals three anchors will be used, lowpass filtered and with reduced stereo image.
Also included is the uncoded original signal, once as open and once as hidden reference.

#	Type	Specification	Channel type
1	Anchor	3.5 kHz Lowpass	Mono
2	Anchor	7.0 kHz Lowpass	Mono
3	Anchor	3.5 kHz Lowpass significantly reduced stereo image (12dB attenuated side channel)	Stereo
5	Anchor	7.0 kHz Lowpass significantly reduced stereo image (12 dB attenuated side channel)	Stereo
6	Anchor	7.0 kHz Lowpass slightly reduced stereo image (6 dB attenuated side channel)	Stereo
7	Hidden Reference	Original signal	Mono/Stereo
8	Open Reference	Original signal	Mono/Stereo

3 References and Conventions

3.1 Reference Documents

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

4 Test Material

The test material will be composed according to the following approach. First, a call will be sent out for test material (2-channels, sampled at 48 kHz) according to a number of generic audio signal categories. Then, an independent selection entity < tbd> will identify a number of items, which are representative for the application scenarios, out of each category to be used in the experiments.

8 different sets of material covering 24 items according to the list below will be used. Four for mono, four for stereo.

Original material will always be in stereo, for mono experiments it will be downmixed following the procedures described in section 8 (processing).

The selection entity will identify based on uncoded material 8 sets of material sampled at 48 kHz out of the following generic audio signal categories:

- Music with sub-categories (8 items)
  - Single Instrument: e.g. piano, violin, clarinet
  - Vocal: e.g.: a cappella, solo singer
  - Choir
  - Pop
  - Classical (orchestra)
  - Other
- Speech over music with sub-categories (4 items)
  - News trailer
  - Advertisement
  - Film sound track
  - Other
  -
- Speech between music with sub-categories (4 items)
  - Speech followed by music
  - Music followed by speech
  - Jingle between speech
  - Speech between jingles
  - Advertisement
  - Other
- Speech with sub-categories (8 items)
  - Clean
  - Noisy: car/street/babble/sports event

For the items out of the speech category tested in the stereo sub-experiments, the following stereo image categories apply:

- Two talkers at different locations
- Moving talker, fixed background
- Fixed talker, moving background

Material with typical stereo images should be well represented in the stereo experiments.

The selected number of items out of each sub-category shall be as balanced as possible. The items containing speech or vocal elements shall be balanced in speaker gender within each sub-category. The speech items shall be in various languages.

The approximate lengths in time of the speech and speech over music items will be 5 s. These items shall comprise one sentence of speech and not contain speech pauses.
The music and speech between music items shall not exceed a length of 10 s.

The complete set of material will be split and equally distributed to the sub-experiments.

4.1 Material selection

The selection entity is/are some independent organization(s) agreed by SA4.

Proposed material has to be submitted to the selection entity after submission date of the candidate codecs and before the material submission deadline according to the schedule.
The selection entity will make the material selection according to the material selection criteria specified in the above section 4.

All 3GPP members are invited to submit test material to host lab. The submitting organization shall assign the items to the above-mentioned content main- and sub-categories. The host lab will blind the material and provide it to the material selection entity.
This ensures the selection is based on items whose origin is not revealed to the selection entity.
The host lab will further maintain and report to SA4 a list indicating the number of proposed items per submitting organization.
In case the submitted material is insufficient/inadequate to conduct the tests, the selection entity will add the missing test items.
The selection entity will provide SA4 with a report about the selection process.

4.2 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 be made with four sound items, one out of each audio signal category. These items will be identified by the selection entity and shall not be re-used in the blind grading phase. The training phase shall be executed as a separate short MUSHRA session.

5 Information relevant to all Experiments

5.1 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.

5.2 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), a number of hidden anchors, the conditions of the codecs under test as well as the reference conditions with which the codecs under test are to be compared.

5.3 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 pattern and its application are given in section 8 (Processing). .

5.4 Training phase

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

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

5.5 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 a more reliable result more quickly than non-experienced listeners.

5.5.1 Screening of subjects

There is sometimes a reason for introducing a rejection technique either before (pre-screening) or after (post-screening) the real test. In some cases both types of rejections might be used. Here, rejection is a process where all judgements from a particular subject are omitted.
Any type of rejection technique, not carefully analysed and applied, may lead to a biased result. It is thus extremely important that, whenever elimination of data has been made, the test report clearly describes the criterion applied.

5.5.1.1 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.

5.5.1.2 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 gradings;
- 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 artefacts, 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.

6 Experimental Block A

6.1 Introduction

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

The experimental block covers four experiments:

- 14 kbps, mono, use case A (PSS)
- 18 kbps, stereo, use case A (PSS)
- 24 kbps, mono, use case A (PSS)
- 24 kbps, stereo, use case A (PSS)

- Each experiment comprises 2 sub-experiments, which are equally designed but carried out with set a, b or c, respectively, of the test material.

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 5. Therefore Listening Laboratories should use the information in Section 5 in conjunction with the information given in this section.

6.2 Test Conditions

Tables 6-2 (a) - (d) provide an overview of the conditions applicable to experimental block A.

Table 6-2 (a): Conditions and factors for Experiment A1 (14 kbps, mono, use case A (PSS))

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Mono
Codec references		
Codec references	2	AAC@14kbps and AMR-WB@14.25 kbps
Input sampling rate		AAC: 48 kHz, AMR-WB: 16 kHz
Number of input channels	1	Mono
Number of output channels	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		48 kHz
Number of input channels	2	Stereo 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
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
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 6-2 (b): Conditions and factors for Experiment A2 (18 kbps, stereo, use case A (PSS))

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Stereo
Codec references		
Codec references	2	AAC@18kbps and AMR-WB@18.25 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Number of output channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	3	7 kHz low-pass filtered original signal with reduced stereo image: 6 dB and 12 dB attenuated side channel 3.5 kHz low-pass filtered original signal with reduced stereo image: 12 dB attenuated side channel
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	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 6-2 (c): Conditions and factors for Experiment A3 (24 kbps, mono, use case A (PSS))

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Mono
Codec references		
Codec references	2	AAC@24kbps and AMR-WB@23.85 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	1	Mono
Number of output channels	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		48 kHz
Number of input channels	2	Stereo 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
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
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 6-2 (d): Conditions and factors for Experiment A4 (24 kbps, stereo, use case A (PSS))

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	No errors
Mono/Stereo	1	Stereo
Codec references		
Codec references	2	AAC@24kbps and AMR-WB@23.85 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Number of output channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	3	7 kHz low-pass filtered original signal with reduced stereo image: 6 dB and 12 dB attenuated side channel 3.5 kHz low-pass filtered original signal with reduced stereo image: 12 dB attenuated side channel
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	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)

6.3 Material

See section 4.

6.4 Experimental Design

See section 5.2

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 8.

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 blind grading phase in each of the sub-experiments this accounts to
12 * 8 * 4 * 2 * 7.5s = 1.6 hours per subject (mono tests) or, respectively,
12 * 9 * 4 * 2 * 7.5s = 1.8 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 (mono tests) or, respectively,
4 * 9 * 4 * 1 * 7.5s = 18 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 Experimental Block B

7.1 Introduction

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

The experimental block covers four experiments:

- 14 kbps, mono, use case B (MMS), 16 kHz sampling rate
- 18 kbps, stereo, use case B (MMS),
- 14 kbps, mono, use case A (PSS), 3% FER
- 24 kbps, stereo, use case A (PSS), 3% FER

Each experiment comprises 2 sub-experiments, which are equally designed but carried out with set a, b or c, respectively, of the test material.

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 5. Therefore Listening Laboratories should use the information in Section 5 in conjunction with the information given in this section.

7.2 Test Conditions

Tables 7-2 (a) - (d) provide an overview of the conditions applicable to experimental block B.

Table 7-2 (a): Conditions and factors for Experiment B1 (14 kbps, mono, use case B (MMS), 16 kHz sampling rate)

Main Codec Conditions		
Candidates	3	
Use case	1	B (MMS)
Error Conditions	1	No errors
Mono/Stereo	1	Mono
Codec references		
Codec references	2	AAC@14kbps and AMR-WB@14.25 kbps
Input sampling rate		16 kHz
Number of input channels	1	Mono
Number of output channels	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
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
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 7-2 (b): Conditions and factors for Experiment B2 (18 kbps, stereo, use case B (MMS))

Main Codec Conditions		
Candidates	3	
Use case	1	B (MMS)
Error Conditions	1	No errors
Mono/Stereo	1	Mono
Codec references		
Codec references	2	AAC@18kbps and AMR-WB@18.25 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Number of output channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	3	7 kHz low-pass filtered original signal with reduced stereo image: 6 dB and 12 dB attenuated side channel 3.5 kHz low-pass filtered original signal with reduced stereo image: 12 dB attenuated side channel
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	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 7-2 (c): Conditions and factors for Experiment B3 (14 kbps, mono, use case A (PSS), 3% FER)

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	3 % FER
Mono/Stereo	1	Mono
Codec references		
Codec references	2	AAC@14kbps and AMR-WB@14.25 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	1	Mono
Number of output channels	1	Mono
Error Conditions of codec references		AMR-WB: 3% FER AAC: 3% FER
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	1	3% FER
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	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
Sub-experiments	2	The experiment is done in 2 sub-experiments with different test material
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 7-2 (d): Conditions and factors for Experiment B4 (24 kbps, stereo, use case A (PSS), 3% FER)

Main Codec Conditions		
Candidates	3	
Use case	1	A (PSS)
Error Conditions	1	3% FER
Mono/Stereo	1	Stereo
Codec references		
Codec references	2	AAC@24kbps and AMR-WB@23.85 kbps
Input sampling rate		AAC: 48 kHz AMR-WB: 16 kHz
Number of input channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Number of output channels	2/1	AAC: 2 (stereo), AMR-WB (mono)
Error Conditions of codec references		AMR-WB: 3% FER AAC: 3% FER
Other references		
Open Reference	1	Original signal
Hidden Reference	1	Original signal
Anchors	3	7 kHz low-pass filtered original signal with reduced stereo image: 6 dB and 12 dB attenuated side channel 3.5 kHz low-pass filtered original signal with reduced stereo image: 12 dB attenuated side channel
Common Conditions		
Stimulus type		Sound item
Radio Channels	1	3% FER
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	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)

7.3 Material

See section 4.

7.4 Experimental Design

See section 5.2

7.5 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.

7.6 Processing

Processing is specified in section 8.

7.7 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 blind grading phase in each of the sub-experiments this accounts to
12 * 8 * 4 * 2 * 7.5s = 1.6 hours per subject (mono tests) or, respectively,
12 * 9 * 4 * 2 * 7.5s = 1.8 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 (mono tests) or, respectively,
4 * 9 * 4 * 1 * 7.5s = 18 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.

7.8 Votes Per Condition

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

7.9 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.
The statistical significance of the results obtained in the sub-experiments with frame erasures may suffer from the effect that – due to different frame sizes of the candidate codecs – the frame erasures become effective at different points in time. The subjective relevancy of the frame erasure will thus not only depend on the frame error mitigation procedure of the codec candidates but also on the signal characteristic of the sound item at the erased frames.
In order to statistically mitigate this latter effect as much as possible there will be a randomization of the error patterns across the various listeners, the two sub-experiments B3 and B4 and the various listening labs. I.e. for each individual listening different error patterns will be used.

8 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, the complete material 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.

8.1 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.

8.1.1 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 tbd.

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

8.1.2 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. The order in the concatenation is such that the training material will precede the test material. 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.

8.2 Main processing

8.2.1 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 “resamp-audio” 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> = [X] 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.

8.2.2 Candidate and reference codec processing

The concatenated material file will be created by the hostlab. After processing, the encoded material as well as the concatenated original files will be delivered to candidates for cross-checking. The candidates 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 candidate codec descriptions. The analogue processing procedure will be applied to the reference codecs. For the AMR-WB reference codec, prior to actual codec processing the concatenated material is P.341 filtered. This filtering operation is performed using the *filter* program from the ITU-T STL (STL 2000):

Source:	ITU-T
Location:	STL sub-directory: fir
Format:	C source code
Program:	*filter*

To produce a P.341 filtered file use:

```
filter P341 input16 output16 320
```

8.2.2.1 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 volunteering organization <tbd>.

8.3 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.

8.3.1 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.

8.3.2 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 $2N$ 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 10 ms. I.e. N=0.01 * 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

8.4 Blinding of material

The purpose of blinding is to unveil the identity of the different codec, reference and anchor conditions. The blinding is done using a script that maps the de-concatenated items onto files with names unveiling the actual condition. The mapping shall vary across the different items.

The blinding of the material is made by the host lab, based on a script created by the host lab. The host lab is responsible to keep the mapping confidential.

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 blind 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 make a short test similar to the one you will make in the blind grading phase of the test.

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

2 Blind grading phase

The purpose of the blind grading phase is to invite you to assign your grades using the quality scale. Your grades should reflect your subjective judgement 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.

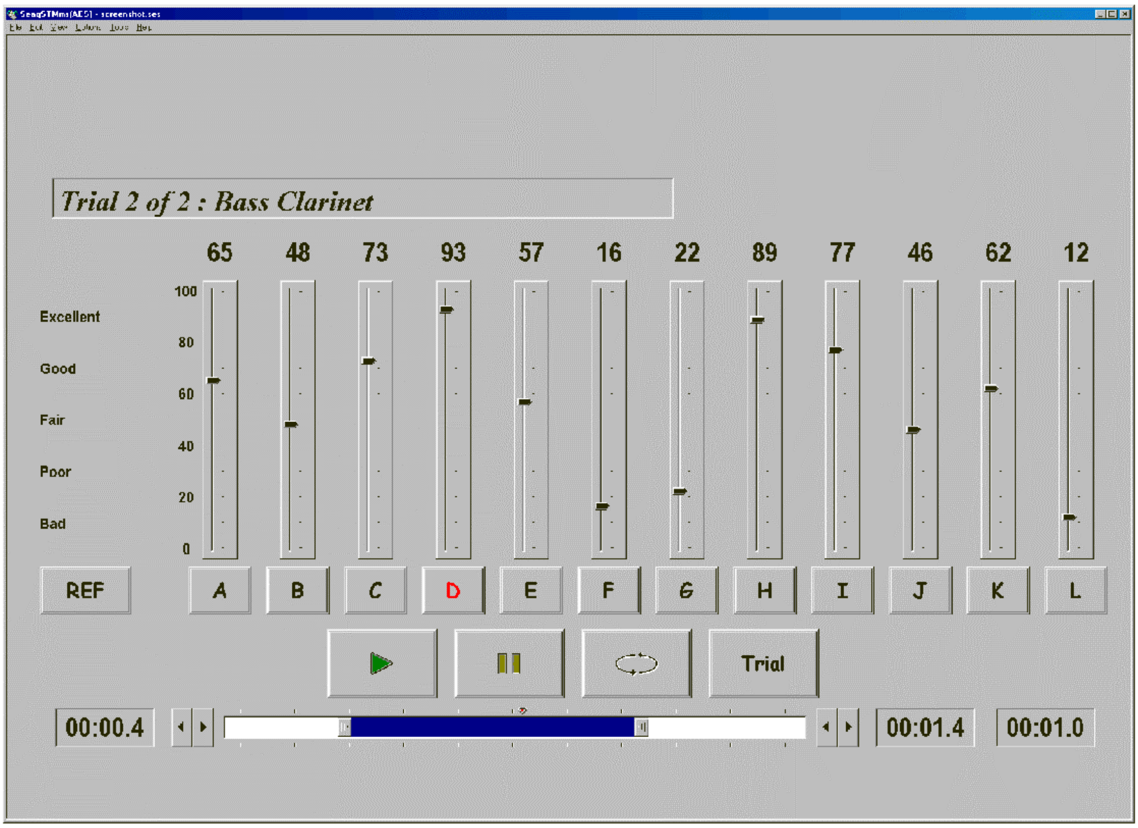
You will use the quality scale as given in Fig. 1 when assigning your grades.

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

```
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 and blinded 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

1. Unprocessed original signals:

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

where

- <main_category> is one out of {m, som, sbm, s} (for music, speech over music, speech between music, speech),
- <sub_category> is one out of
  - {si, vo, ch, po, cl, ot} for music main category (single instrument, vocal, choir, pop, classical, other)
  - {ne, ad, fi, ot} for speech over music main category (news trailer, advertisement, film sound track, other)
  - {sm, ms, js, sj, as, ot} for speech between music main category, and
  - {cl, no} for speech main category (clean, noisy)
- <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.
- <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

- <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 after blinding:

<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.