**Summary**

This document describes the results of Experiment3 of the 14kHz Low-Complexity Audio Coding Algorithm at 24, 32 and 48 kbps Extension to ITU-T G.722.1 Subjective Characterisation Tests performed by France Telecom. In the following, this codec will be called, for convenience, "G.722.1C". The test was performed according to the ITU-T Characterisation Test Plan using the processed material provided by France Telecom.

In summary, in experiment 3, G.722.1C at 24, 32 and 48 kbps is scored better than AAC-LD at the same bit rate.

# Contributors

This document has been reviewed by the ITU-T Q7/12 experts group under the chairmanship of the two Q7/12 Rapporteurs in liaison with ITU-T Q10/16.

**Q7/12 Rapporteurs**:

|  |  |
| --- | --- |
| Catherine Quinquis  France Telecom  France | Tel: +33 2 96 05 14 93  Fax: +33 2 96 05 3530  Email: [catherine.quinquis@francetelecom.com](mailto:catherine.quinquis@francetelecom.com) |
| Paolo Usai  ETSI  France | Tel: +33 4 92 94 42 36  Fax: +33 4 92 94 52 06  Email: [paolo.usai@etsi.org](mailto:paolo.usai@etsi.org) |

# Experiment 3

France Telecom tested G.722.1C for sub-experiment 3a, 3b and 3c of the Characterisation Test according to the specifications in the Characterisation Test Plan. France Telecom provided the processed material used in this experiment.

The codec performance was assessed for music and mixed content. The Experiment was performed using the Mushra method with P.341extension-weighted speech.

# Test process

## Test method

The methodology MUSHRA was used for those five quality test. MUSHRA stands for MUlti Stimuli with Hidden Reference and Anchor points. This is a method dedicated to the assessment of intermediate quality.

It has been recommended at the ITU-R under the name BS.1534[[1]](#footnote-2).This was developed in 1999 by the EBU Project Group B/AIM in collaboration with the ITU-R Working Party 6Q.

An important feature of this method is the inclusion of the hidden reference and bandwidth limited anchor signals. For those three mentioned tests, anchor points were the band-limited (7 and 10 kHz) reference signal.

## Training phase

Each listener had a period of training, in order to get familiar with the test methodology, the use of the interface software and with the kind of quality they have to assess. This was as well an opportunity to adjust the restitution level that then remained constant during the test phase.

As there were 3 tests, each of them was preceded by a training phase that each listener was asked to perform.

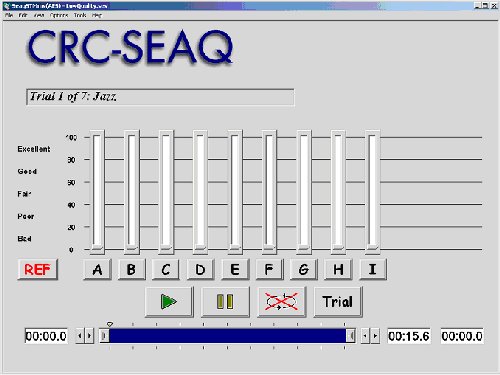
Each training session contained 2 audio items that were different from the audio excerpts played in the tests.

## User Interface

The MUSHRA method has the advantage of displaying all stimuli for one test item at a given bit-rate at the same time. The subjects were therefore able to carry out any comparison between them directly as well as to assess the quality comparing to the one of the explicit reference signal.

Implementation of MUSHRA user interface from CRC (SEAQ) was used in those tests. A screenshot of one implementation of the user interface is shown in figure 1. The buttons represent all the configurations/codecs under test including the hidden reference and both the anchor signals, and the reference, which is specially displayed on the left as "REF". Above each button, with the exception of the "REF" one, a slider is used to grade the quality of the test item according to the continuous quality scale.

For each of the test items, the signals under test were randomly assigned, with a different assignment for each subject. In addition, the test items were randomised for each subject within a session to avoid sequential effects. The session files were prepared by the host lab. There was one session file per listener.



**Figure 1 :** MUSHRA Software

## The Listening Panel

The listening panel consisted of 20 subjects for each sub-experiment, most of them experienced in audio but not only professionally involved. All of them were respectful regarding the listening instructions.

## Tests duration

As mentioned above there were 3 different tests, all preceded by a training period.

The training phase took about half an hour. This time was also used to describe the listening instructions and answer listeners' questions if any.

Then, one test took approximately 1 hour and a half (depending on listeners), including breaks. Every 20 minutes, the listener was asked to rest a bit by breathing some fresh air.

## Listening conditions

The tests were performed over open-back, circum-aural headphone and an amplifier TASCAM MH-40MKII. The subjects had the possibility to set the reproduction level individually before they started the actual test (during the training phase). The subjects were then restricted from changing the reproduction level during the test.

The test items were stored on a Windows 2k workstation. The digital sound was played through the PC board Digigram VX 222 and converted by 24 bits DAC (3Dlab DAC 2000).

The tests were run in an acoustically neutral room dedicated to such tests.

# Test results

# Experiment 3a

The results of Experiment 3a are in table 1

**Table 1** – Mean scores for Experiment 3a

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Conditions | Averaged  mean scores on music items | SD  music items | Averaged  mean scores on mixed content items | SD  mixed content items | Averaged  mean score | SD |
| aac-ld at 24kbps | 34.15 | 25.88 | 22.77 | 16.23 | 28.46 | 22.28 |
| CuT at 24 kbps | 67.96 | 23.18 | 56.37 | 26.70 | 62.17 | 25.60 |
| enh.aacPlus at 24 kbps | 72.76 | 25.39 | 58.84 | 24.86 | 65.80 | 26.02 |
| ext. amrwb at 24kbps | 73.27 | 22.55 | 72.25 | 19.76 | 72.76 | 21.16 |
| hidden ref | 97.39 | 8.41 | 98.73 | 4.41 | 98.06 | 6.73 |
| lp10 | 68.77 | 24.12 | 73.24 | 22.35 | 71.01 | 23.30 |
| lp7 | 38.23 | 26.71 | 33.33 | 17.44 | 35.78 | 22.64 |

# 



# Experiment 3b

The results of Experiment 3b are in table 2

**Table 2** – Mean scores for Experiment 3b

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Conditions | Averaged  mean scores on music items | SD  music items | Averaged  mean scores on mixed content items | SD  mixed content items | Averaged  mean score | SD |
| aac-ld at 32kbps | 54.43 | 28.43 | 38.66 | 23.37 | 46.55 | 27.14 |
| CuT at 32 kbps | 72.29 | 26.19 | 58.08 | 29.43 | 65.19 | 28.68 |
| enh. aacPlus at 32kbps | 81.55 | 20.67 | 63.59 | 26.69 | 72.57 | 25.46 |
| ext. Amrwb at 32 kbps | 77 | 22.91 | 75 | 24.19 | 76 | 23.54 |
| hidden ref | 96.11 | 10.08 | 97.17 | 5.86 | 96.98 | 8.27 |
| lp10 | 58.95 | 26.28 | 60.03 | 23.72 | 59.49 | 24.97 |
| lp7 | 32.60 | 25.24 | 30.41 | 18.12 | 31.51 | 21.94 |

# 

# 

# Experiment 3c

The results of Experiment 3c are in table 3

**Table 3** – Mean scores for Experiment 3c

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Conditions | Averaged  mean scores on music items | SD  music items | Averaged  mean scores on mixed content items | SD  mixed content items | Averaged  mean score | SD |
| aac-ld at 48 kbps | 67.10 | 22.39 | 51.85 | 22.60 | 59.48 | 23.71 |
| CuT at 48 kbps | 85.92 | 20.57 | 79.43 | 20.63 | 82.68 | 20.80 |
| aac-ld at 64kbps | 65.28 | 23.50 | 50.40 | 21.61 | 57.84 | 23.72 |
| hidden ref | 96.64 | 9.08 | 97.80 | 6.19 | 97.22 | 7.77 |
| lp10 | 62.62 | 24.19 | 65.23 | 23.80 | 63.93 | 23.97 |
| lp7 | 34.63 | 21.89 | 32.11 | 16.94 | 33.37 | 19.57 |

# 



1. ITU-R Recommendation BS.1534 (June 2001)/ Method for the subjective assessment of intermediate quality level of coding systems. [↑](#footnote-ref-2)