Source: Coding Technologies

**Title: Evaluation of codec behaviour with special input signals**

**Agenda Item: 7, 12.2**

**Document for: Approval**

1. Introduction

At the SA4 #36 meeting, Coding Technologies volunteered to provide input for the characterization item “Content dependency”, as per Tdoc S4-050623. This document contains the input that Coding Technologies was able to collect by the submission deadline. The results collected in this document will be included into the TR 26.936.

1. Description of the task performed

Due to the conviction that good and consistent performance for music signals over a wide variety of material is essential for an audio codec, and further due to the high commercial relevance of music services for the mobile industry, Coding Technologies decided to concentrate on the behaviour of the codecs using music signals spanning a large range of signals as well as certain specific characteristics. Those signal characteristics were chosen since they are known to be critical for audio codecs and since they were not sufficiently covered by the signals used in the selection and characterization tests so far. By verifying the codec behaviour for these kinds of critical signals, the aim was to establish if any systematic problems are present in the codecs, that have not been discovered during the tests performed so far in the selection and/or characterization tests. In particular we selected material which had the following characteristics (names in brackets are the titles of the items used in the listening test):

* 1. complex stereo image (choir, EarlyWarn, Emotions, sm03, vegainstru)
  2. low correlation between left and right channel (IFeelFine, JukeBox, ManOn, Paperback. SonFather)
  3. transients (beethoven, higher, OceanOf, tas, xylophone),
  4. important tonal components (Because, C\_Mc05, ccr, Rory5, te27)

Work was performed in a three-step approach. In a first step, a team of research engineers identified samples in our audio database which matched the above mentioned characteristics, the loudness level of the selected originals was aligned in-between each other. In a second step, an informal listening exercise was performed to determine whether any systematic effects could be observed for the codecs. Based on the fact that such systematic effects appeared to be present, we conducted a formal listening test on a subset of the previously identified material to confirm the impression from the informal listening.

Detailed information on the origins of the test-items is given in the table below. It should be noted that all signals were taken directly from a CD or from existing test sets, none of the items was pre-processed (except for loudness-level alignment) or artificially generated by Coding Technologies to create a specific characteristic.

|  |  |  |
| --- | --- | --- |
| **Label** | **Title** | **Origin** |
| choir | Ich hatte viel Bekümmernis | Kammerchor Nürnberg |
| EarlyWarn | Early Warning | Rory Gallagher, Photo-finish ( Digitally Remastered ) |
| Emotions | Emotions | Mariah Carey, Emotions |
| Sm03 | Sm03 (“Plucked Strings”) | MPEG test-set |
| vegainstru | Tom’s diner reprise (instrumental) | Suzanne Vega, Solitude Standing |
| IFeelFine | I feel fine | The Beatles, 1 |
| JukeBox | Juke Box Annie | Rory Gallagher, Photo-finish ( Digitally Remastered ) |
| ManOn | Man on the rock | Astral Doors, Of the Son and the father |
| Paperback | Paperback Writer | The Beatles, 1 |
| SonFather | Of the son and the father | Astral Doors, Of the Son and the father |
| beethoven | Streichquartett op. 18 Nr. 1 | Ludwig van Beethoven (Audio Hörkurs 05/2005) |
| higher | Higher | Ezio, Higher |
| OceanOf | Ocean of Sand | Astral Doors, Of the Son and the father |
| tas, | Of Snails and Flies | Teasing a Sphinx, Allegory |
| xylophone | Xylophone | Audio Hörkurs 05/2005 |
| Because | Because | The Beatles, Abbey Road |
| C\_Mc05 | C\_Mc05 | NAB test set |
| Ccr | Creedence Clearwater Revival |  |
| Rory5 | I'll admit you're gone | Rory Gallagher, Calling Card |
| te27 | Triangles | MPEG-2 verfication item (source SQAM) |

1. Description of the listening tests

Two MUSHRA tests were performed. Each test was conducted by two test labs. Coding Technologies conducted both tests (with 9 resp. 10 listeners), Fraunhofer (8 listeners) and Philips (8 listeners) served as the mirror lab for one of the tests each.

Each test included the hidden reference and the usual anchor signals, as well as the 2 codecs under test at three different bitrates. In particular, these bitrates were 16 kbps stereo, 24 kbps stereo and 32 kbps stereo. In addition, AMR-WB+ was further tested at 48 kbps stereo.

Each test used 12 input signals. Test “AB” included five samples from each of characteristic a) and b) plus two music samples taken from the characterization test set (m\_ot\_x\_1, m\_p\_x\_1). Test “CD” included five samples from each of characteristic c) and d) plus two music samples taken from the characterization test set (same as in test “AB”). The objective of including samples from the characterization test was to verify whether the listening test yields comparable results to the selection test for these items, in order to provide a further anchor for interpretation of the results on the special test signals.

Apart from scoring the different codecs for different items and bitrates in the MUSHRA test, the listeners were also requested to make notes motivating the scores given, so that a better understanding could be gained when evaluating the results. Naturally, both the scores and comments are given in a blind fashion since the listeners do not know which codec he or she is listening to.

1. Listening test results

The average listening test results can be seen in the following four diagrams. Diagram 1 and 2 show the average performance across items for the two test sites for test “AB”, diagrams 3 and 4 show the average performance across items for the two test sites for test “CD”. Enhanced aacPlus is referred to as “AP”, AMR-WB+ is referred to as WB, and the two-digit number denotes the bitrate in kbps.



*Diagram 1: Average scores for test AB, test site CT*



*Diagram 2: Average scores for test AB, test site FhG*

It is clear that for the test-set AB (i.e. signals with complex stereo image and low correlation between the two channels), the results are similar between the two test-sites. Both codecs exhibit increasing quality with increasing bitrate. Enhanced aacPlus shows a particularly large quality increase between 16 kbit/s and 24 kbit/s and performs better than AMR-WB+ at all tested bitrates. It should be noted that Enhanced aacPlus at 32 kbit/s also outperforms AMR-WB+ at 48 kbit/s with these input signals.

Studying the comments accompanying the scores given, it becomes obvious that the AMR-WB+ has significant problems with instability of the stereo image with clearly audible switching between a mono and a stereo impression.



*Diagram 3: Average scores for test CD, test site CT*



*Diagram 4: Average scores for test CD, test site Philips*

For the CD test-set (i.e. signals with transients and important tonal components), the results show similar characteristics as for the AB test. Again, both codecs show increased quality with increasing bitrate and the overall quality of Enhanced aacPlus is judged superior to that of the ARM-WB+ codec.

Considering the comments given by the listeners again, it seems that the main issue for the AMR-WB+ codec is bad transient response for the class c) items. Furthermore, for the class d) items of test CD, the AMR-WB+ codec shows severe blocking artefacts that introduces artificial periodicities during tonal parts of the signals.

Below is a summary of what words that were used to describe the artifacts for the two codecs at the different bitrates.

|  |  |
| --- | --- |
| Codec and bitrate | Comments |
| AMR-WBplus 16 kbit/s | wobbling, distortion, bubbly, blocky, unstable, metallic |
| AMR-WBplus 24 kbit/s | wobbling, distortion, blocky, metallic, unstable, |
| AMR-WBplus 32 kbit/s | unstable stereo image, wobbling, distortion, ghost voice, blocky, |
| AMR-WBplus 48 kbit/s | narrow stereo image, unstable, missing presence, distortion |
| Enhanced aacPlus 16 kbit/s | metallic, sharp, bubbly, diffuse, |
| Enhanced aacPlus 24 kbit/s | diffuse ,slightly metallic, slightly narrow stereo image |
| Enhanced aacPlus 32 kbit/s | diffuse ,missing presence |

In a test where the behavior for special input signals is of major interest, the per-item scores are of course specifically interesting. Figures 5 to12 show the full per-item scores with all codecs and anchors present (2 graphs per test to fit paper size). Figures 13 to 20 show excerpts of the per-item scores, focusing on a subset of the tested codecs each, for enhanced readability.



*Diagram 5: Per-item scores for test AB, test site CT (part 1)*



*Diagram 6: Per-item scores for test AB, test site CT (part 2)*



*Diagram 7: Per-item scores for test AB, test site FhG (part 1)*



*Diagram 8: Per-item scores for test AB, test site FhG (part 2)*



*Diagram 9: Per-item scores for test CD, test site CT (part 1)*



*Diagram 10: Per-item scores for test CD, test site CT (part 2)*



*Diagram 11: Per-item scores for test CD, test site Philips (part 1)*



*Diagram 12: Per-item scores for test CD, test site Philips (part 2)*



*Diagram 13*



*Diagram 14*



*Diagram 15*



*Diagram 16*



*Diagram 17*



*Diagram 18*



*Diagram 19*



*Diagram 20*

Looking at the per-item results for the signal characteristics a) and b) (diagrams 5 to 8 and diagrams 13 to 16) we can see for most items a significant degradation in subjective quality relative to the two reference tracks from the characterization test for AMR-WB+. We can also observe a similar effect for Enhanced aacPlus, but only on very few items, in particular for “choir”. The degradation appears to be more pronounced at higher bitrates, however this may be related to floor effects.

When looking at the diagrams 9 to 12 and 17 to 20 respectively (signal characteristics c) and d) ), we again recognize a clear degradation in subjective quality relative to the two reference tracks from the characterization test for AMR-WB+ for most of the test items. We can also observe a similar effect for Enhanced aacPlus, but only on very few items, in particular for “te27”.

1. Conclusions

The tests which have been presented in this document focus on real-world music signals with special characteristics, some of which would be considered critical. The tested material is largely trying to close gaps in the signal characteristics used in the selection and characterization test so far. The goal of the test was to determine whether such material reveals any systematic shortcomings in any of the two codecs as well as the performance consistency across a wider range of critical music samples.

The tests reveal several aspects which appear to be systematic and lead to a significant performance degradation for the ARM-WB+ codec. Relative to average material such as that used in the selection and characterization tests the performance degradation has to be considered significant, leading to an overall inconsistent codec behaviour.

In particular, the AMR-WB+ codec showed systematic degradation for signals with complex stereo images, signals with a low correlation between the two channels, strong transients and pronounced tonal components.