**Source: Dynastat[[1]](#footnote-2)1**

**Title: Global Analysis Laboratory Report for Phase 2 of the 3GPP Audio Codec Characterization Test for PSS-MMS-MBMS Applications**

**Agenda item: 8, 11.1**

# Introduction

This document comprises the final report for the Phase 2 activities of the Global Analysis Laboratory for the Characterization of the 3GPP Audio Codecs. It summarizes the results and analyses from Phase 2 of the Characterization Test. Phase 2 included four listening tests evaluating the subjective performance of the two audio codecs under conditions of Packet Loss.

# Organization of the Characterization Test

The Characterization Test Plan [1] specified the subjective listening tests to characterize the performance of the two audio codecs, *Enhanced aacPlus* (Eaac+) and *Extended AMR-WB* (AMR-WB+), selected by 3GPP for standardization for PSS, MMS, and MBMS applications. The test plan specified that the subjective tests use the “Multiple Stimulus with Hidden Reference and Anchors” or MUSHRA test method [2] for the subjective assessment of intermediate audio quality. The MUSHRA experiments were subdivided into two phases of testing. Phase 1 included two MUSHRA tests characterizing the performance of the two audio codecs across bit rates. Each of the two Phase 1 MUSHRA experiments was conducted in two Listening Labs. The results of the Phase 1 tests were presented in an earlier report to 3GPP/SA4 [3]. For the Phase 2 testing, the test plan describes four MUSHRA tests to characterize the performance of the two audio codecs under conditions of Packet Loss Rate (PLR).

Experiments 2-1 and 2-2 characterized the two audio codecs for the Enhanced GPRS (EGPRS) application -- Exp.2-1 for the Mono mode and Exp.2-2 for the Stereo mode. In both experiments the two audio codecs were tested across four PLR’s, 0%, 1%, 6%, and 10%. In Exp.2-1, AMR-WB+ was operating at 16kbps and Eaac+ at 20kbps. In Exp.2-2, both codecs were operating at 24kbps.

Experiments 2-3 and 2-4 characterized the two audio codecs for the UMTS Terrestrial Radio Access Network (UTRAN) application -- Exp.2-3 for Stereo/lower bit-rate and Exp.2-4 for the Stereo/higher bit-rate. In both experiments the two audio codecs were tested across three PLR’s, 0%, 1%, and 5%. In Exp.2-3, AMR-WB+ was operating at 20kbps and Eaac+ at 32kbps. In Exp.2-4, both codecs were operating at 40kbps.

In addition to the test-conditions, all four experiments also included the four anchor and reference conditions prescribed by the MUSHRA standard:

* Anchor-conditions
  + 3.5k low-pass anchor
  + 7.0k low-pass anchor
* Reference-conditions
  + Open Reference – Original source signal (presented but not evaluated)
  + Hidden Reference – Original source signal (same condition as the Open Reference)

For the Stereo experiments (Exps. 2-2, 2-3, and 2-4), the two anchor conditions were further degraded with a Reduced Stereo Image (RSI), 6dB RSI for the 7.0kHzLP anchor and 12dB RSI for the 3.5kHzLP anchor. Table 1 presents a summary of the experiments and conditions designed for the Phase 2 tests.

Table 1. Test and Reference Conditions in the Phase 2 MUSHRA Experiments.



# Methods and procedures

## Host Labs

The test plan allocated responsibility for Host Lab (HL) processing and cross-checking to the proponent companies for the two audio codecs. Coding Technology (Eaac+) processed the audio test conditions for Exps. 2-1 and 2-2 and cross-checked the conditions for Exps. 2-3 and 2-4. Ericsson/Nokia (AMR-WB+) processed the conditions for Exps. 2-3 and 2-4 and cross-checked the conditions for Exps. 2-1 and 2-2. The HL’s also delivered the processed audio files to the LL’s for testing.

## Error Patterns and Error Conditions

The test plan specified that the error conditions in the Phase 2 tests would be processed according to AMR-WB+ performance requirements [5] and PSS/MMS Audio Codec Selection Design Constraints and Performance Requirements – Version 2.0. Ericsson provided the error patterns for Exps. 2-1 and 2-2 and Qualcomm provided the error patterns for Exps. 2-3 and 2-4.

## Listening Conditions

The test plan specified the methods and procedures for conducting the listening tests. Each test used the MUSHRA methodology designed for the subjective assessment of intermediate audio quality. The test plan also specified the listening conditions to be employed by the LL. For all experiments, subjects should be seated in a quiet environment; 30dBA Hoth Spectrum [4] measured at the head position of the subject. The test stimuli will be presented to the subjects for binaural listening using closed-back/supra-aural headphones or open-back/circum-aural headphones.

Each of the four MUSHRA tests was conducted by a different LL. Each LL presented a report detailing the methods and procedures used for conducting their MUSHRA test. Table 2 summarizes the LL reports and indicates the reference to those reports (in brackets). Only one LL reported a deviation from test plan specifications in conducting their MUSHRA test. In Exp. 2-4 a script error resulted in the same signal being presented to all listeners for the AMR-WB+, 5%PLR condition whereas the test plan specified different signals for each subject.

Table 2. Summary of Listening Labs and Subjects for the MUSHRA tests.

:



As specified in the test plan, each of the four MUSHRA tests involved the same 12 audio items and each item was processed through each test and reference condition involved in the test. The audio items were selected to represent three classes of Audio Content – Music, Speech, and Mixed Music/Speech Content. The Mixed Content class was further sub-classified into Speech-Over-Music and Speech-Between-Music items. Among the 12 test audio items, there were four items for each of the three classes of Audio Content.

The test plan required each LL to deliver raw voting data for 15 expert listeners. The GAL provided each LL with an Excel spreadsheet for delivery of the raw voting data. Each LL delivered raw MUSHRA voting data (180 votes = 15 listeners x 12 items) to the GAL prior to the deadline prescribed by the test plan.

# Overall Results

While the test plan listed target coding bit-rates for the two audio codecs in each of the four MUSHRA tests, the source coding rates used for processing the test conditions differed somewhat from the target values. Table 3 shows the various bit-rates (Bearer rate, RTP rate, Target coding rate, and Source coding rate) for the two codecs in the four MUSHRA experiments. However, in the tables and figures that follow, the Target coding rate will be used to identify the condition.

Table 3. Codec Bit-Rates for the Four Phase 2 MUSHRA Experiments



Table 4 shows summary results for Exp.2-1, EGPRS in Mono Mode. The results include the MUSHRA Mean and Standard Deviation for each test and reference condition and are based on 180 votes (15 subjects x 12 test items). Figure 1 illustrates the results of Exp.2-1 graphically. The figure shows MUSHRA Means and 95% Confidence Intervals for each test and reference condition involved in the MUSHRA experiment.

Table 4. MUSHRA Results for Exp.2-1 (EGPRS, Mono).





Fig. 1. MUSHRA Results for Exp.2-1.

Table 5 shows summary results for Exp.2-2, EGPRS in Stereo Mode, and Fig. 2 illustrates those results graphically.

Table 5. MUSHRA Results for Exp.2-2 (EGPRS, Stereo).





Fig. 2. MUSHRA Results for Exp.2-2.

Table 6 and Fig.3 show summary results for Exp.2-3, UTRAN in Stereo Mode with lower bit rates.

Table 6. MUSHRA Results for Exp.2-3 (UTRAN, Stereo, lower bit rates).





Fig. 3. MUSHRA Results for Exp.2-3.

Table 7 and Fig.4 show summary results for Exp.2-4, UTRAN in Stereo Mode with higher bit rates

Table 7. MUSHRA Results for Exp.2-4 (UTRAN, Stereo, higher bit rates).





Fig. 4. MUSHRA Results for Exp.2-4.

An initial examination of Figs. 1-4 leads to the following observations:

* All Experiments
  + Performance decreases for both codecs with increases in PLR
* Exp.2-1 – EGPRS – Mono
  + AMR-WB+ at 16k shows similar performance to Eaac+ at 20k for low values of PLR (0% and 1%) but worse performance for higher values of PLR (6% & 10%)
* Exp.2-2 – EGPRS – Stereo
  + AMR-WB+ at 24k shows similar performance to Eaac+ at 24k for PLR of 0% and 1% but worse performance for PLR of 6% & 10%
* Exp.2-3 – UTRAN – Stereo
  + Both codecs show similar pattern of scores across PLR but Eaac+ at 32k shows better performance than AMR-WB+ at 24k
* Exp.2-4 – UTRAN – Stereo
  + Eaac+ at 40k performed better than AMR-WB+ at 40k especially with increases in PLR

It is important to note that since the four tests were conducted by different LL’s, each with a different listening instrument, different subjects, different languages, etc., raw scores should not be compared across experiments. Comparisons are valid and meaningful only on those conditions that are contained within the same MUSHRA experiment.

# Effects of Audio Content

The twelve audio items involved in each experiment represented three classes of Audio Content: *Music* only, *Speech* only, and *Mixed* content – speech and music. Figure 7 shows the MUSHRA results, Means and 95% Confidence Intervals, for Exp.2-1. In the figure MUSHRA results are shown for each test and reference condition for each of three classes of Audio Content.



Fig.7 Results for the Conditions in Exp. 2-1 by Audio Content

Figures 8, 9, and 10 show the MUSHRA results by condition and by Audio Content for Exps. 2-2, 2-3, and 2-4, respectively.



Fig.8 MUSHRA Results for the Conditions in Exp. 2-2 by Audio Content



Fig.9 MUSHRA Results for the Conditions in Exp. 2-3 by Audio Content



Fig.10 MUSHRA Results for the Conditions in Exp. 2-4 by Audio Content

An examination of Figs. 7-10 leads to the following observations:

* All Experiments – there were differences in performance across classes of Audio Content and between the two audio codecs for classes of Audio Content
* Exp.2-1 – EGPRS – Mono
  + The two codecs showed similar performance across PLR and across classes of Audio Content
* Exp.2-2 – EGPRS – Stereo
  + AMR-WB+ shows relatively worse performance for Music Content, especially at lower values of PLR
* Exp.2-3 – UTRAN – Stereo
  + Eaac+ shows relatively worse performance for Speech Content
* Exp.2-4 – UTRAN – Stereo
  + AMR-WB+ shows relatively worse performance for Music Content and relatively better performance for Speech and Mixed Content while Eaac+ shows relatively better performance for Music Content and relatively worse performance for Speech and Mixed Content

# References

[1] Tdoc S4-050440 PSS/MMS/MBMS Audio Codec Characterization Test Plan Version 0.7, May 2005.

[2] EBU Technical recommendation: MUSHRA-EBU Method for Subjective Listening Tests of Intermediate Audio Quality, Doc. B/AIM022, Oct.1999.

[3] Tdoc S4-050428 Global Analysis Laboratory Report for Phase-1 Audio Codec Characterization Test for PSS-MMS-MBMS

[4] ITU-T Recommendation P.800: Methods for subjective determination of transmission quality, August 1996

[5] Tdoc S4-030434 AMR-WB+ Performance Requirements, Version 2.0

[6] Tdoc S4-030448 Fraunhofer IIS (Exp.2-1) Listening laboratory report in the course of the GPP/PSS/MMS/ MBMS audio codec characterization test

[7] Tdoc S4-030449 NTT-AT (Exp.2-2) Report on PSS/MSS Audio Codec Characterization Test

[8] Tdoc S4-030452 Nokia (Exp.2-3) Listening test laboratory report

[9] Tdoc.S4-030454 T-Systems (Exp.2-4) Listening test laboratory report on the 3GPP PSS/MMS/ MBMS audio codec characterization test (Phase 2)

1. 1 Alan Sharpley

   Dynastat, Inc. Email: asharpley@dynastat.com

   6850 Austin Center Blvd., Suite 150 Phone: +1-512-476-4797

   Austin, Texas, USA 78731 FAX: +1-512-472-2883 [↑](#footnote-ref-2)