Evaluation of Different AAC Codec Realizations for Audio Mobile Device based on ARM Architecture

LAZAR BERBAKOV, STEFAN TRAILOVIĆ, DEJAN TODOROVIĆ,
VUKAŠIN RISTIĆ, NIKOLA NENADIĆ, IVA SALOM, GORAN FERENC
Institute Mihailo Pupin
University of Belgrade
Volgina 15, 11060 Belgrade
SERBIA
{lazar.berbakov, stefan.trailovic, vukasin.ristic, nikola.nenadic, iva.salom, goran.ferenc}@pupin.rs
dejan.todorovic@dirigent-acoustics.co.rs
http://www.pupin.rs

Abstract: - In this paper, we present objective and subjective evaluation of AAC audio codec implementations for ARM based audio mobile device. Selected audio material was coded and decoded using implementations provided by four different vendors and then compared to the original audio recordings by using EAQUAL software. Afterwards, two implementations with highest scores were evaluated by using subjective A/B/X listening test. Finally, the optimal choice of codec and bitrate was made, based both on subjective quality of the codecs and on other important objective indicators such as processor usage and memory occupation of ARM architecture implementations.

Key-Words: - AAC codec evaluation, Subjective audio test, Objective audio test

1 Introduction

In this paper, we present an evaluation of different AAC audio codec implementations based on objective and subjective quality grades. This evaluation has been performed for the purpose of the development of a battery powered standalone audio recording device aimed to instantly capture inspiration of the musicians and creative people.

This device will have a built-in internal memory in order to record sound, wireless connectivity to iPhone and Android based devices and more significantly - studio grade signal quality. The device will integrate several existing devices in one:

- ***Analog microphone***: Device will include 3.5 mm mini stereo jack headphone output with microphone monitor function. It means that this device will have analogue output that could be connected to an external audio card in order to record sound (e.g. on a PC), like with any other analog microphone.
- ***Recorder***: Internal recorder with memory capacity of 8 GB will provide recording of 16 hours in WAV format or 160 hours in AAC format. The device will integrate USB mass storage interface, so that recorded files will be accessible via host computer.
- ***Playback module***: Besides local monitoring, recorder files playback functionality will be integrated as well. In such a way, musicians will be able to listen audio files they recorded earlier. Furthermore, device will provide the so-called *recording on top* functionality, which means that it will be possible to make a new recording over the track that has already been recorded and currently played back on the headphones.
- ***USB microphone***: USB audio functionality will be one of possible use cases the device will provide, so that it could be used as an USB microphone.
- ***Wireless microphone***: The device will provide real-time wireless streaming to iPhone and Android based devices via Bluetooth 2.1 (Classic Bluetooth) interface. Using an application on a mobile phone, it will be possible to edit, mix and share recorded tracks. Besides, it will be possible to connect up to 3 devices and stream data to the application at the same time.

Golden plated condenser capsule with 18 mV/Pa rated sensitivity will be used as an audio sensor. Its frequency response is in the range from 20 Hz to 20 kHz. In addition, Li-Ion battery with 1000 mAh rated capacity will provide 7 hours of local recording or 5 hours, if Bluetooth streaming is used. Battery will be charged over the micro USB connector, since battery charger will be integrated in the device itself. Finally, the user will be given the option to choose between 16-bit or 24-bit resolutions and 44.1 kHz, 48 kHz and 96 kHz sampling frequencies.

In order to achieve reasonable manufacturing cost of the final device, with the longest possible autonomy and required audio quality, we have paid special attention to the choice of hardware components, their energy consumption, as well as computational complexity and audio quality of used audio codec.

In order to enable real-time audio data transmission, with the given constraint on wireless connection bandwidth, we have chosen AAC codec [1], which was shown to provide excellent audio quality on lower bitrates. Besides, in some previous works [2], it was shown that AAC codec provides better audio quality when compared to popular MP3 codec at the same bitrate.

We have chosen four different AAC codec implementations (from four different vendors) that were ported to the hardware platform of interest for this project. For the codec quality evaluation purposes, a set of high quality audio recording has been prepared so that different categories such as: vocal, solo instrument and orchestra are present. Chosen audio recordings were compressed using AAC encoders with different combinations of bit depth (16/24 bit) and bitrate (96 – 256 kbps). In addition, computational resource consumption of each AAC codec implementation was determined.

Using only subjective tests of audio quality turned out to be impractical due to a large number of codec parameter combinations. In order to overcome this problem, in the first pass we evaluated the objective quality grade (ODG) of each compressed audio recording by comparing it with the original (raw) one, using the software tool EAQUAL (Evaluation of Audio Quality) [3]. This tool, based on ITU-R recommendation BS.1387 [4], provides objective quality grade for compressed audio recordings by modelling physical, psycho-acoustic and cognitive properties of human auditory system.

Next, we have chosen two codec implementations with the highest objective quality grade, and then we have evaluated their performance by carrying out subjective tests. Finally, the optimal codec has been identified by taking into account not only the objective and subjective audio quality grades, but also the computational complexity and the vendor's licensing model.

2 Theoretical background

2.1 AAC (Advanced Audio Coding)

AAC is an audio coding standard for lossy audio compression. It represents a widely used codec that employs the same encoding principles as MP3. Besides, it improves on MP3 in terms of encoding efficiency and audio features types [1], [5], resulting in better audio quality, especially at lower bitrates. AAC supports 48 full bandwidth audio channels in one stream (sampling frequency from 8 to 96 kHz), plus 16 low frequency effects (limited to 120 Hz), up to 16 dialog channels and 16 data streams. Stereo audio signal provides satisfactory to modest quality at 96 kbps in joint stereo mode. Besides, the MPEG-2 audio tests showed that the "transparent" quality for stereo signal is achieved at 128 kbps and for 5.1 audio at 320 kbps.

Similarly to MP3, AAC codec reduces the amount of data necessary to describe the audio signal by removing those elements that could not be heard due to imperfections of human auditory system. Although tests [5] in the late 1990s showed that AAC exhibited better sound quality and transparency than MP3 for files coded at the same bit rate, numerous listening tests performed later have shown that the best encoders for each format are often of similar quality. It was shown that the quality is often dependent on the encoder implementation, even for the same format. Although AAC advantages over MP3 are evident below bitrate of 100 kbps, certain AAC encoders do not employ additional encoding strategies defined for AAC, resulting in slightly lower audio quality than for the best MP3 encoder.

In the sequel, we provide the list of improvements introduced by AAC codec:

- AAC supports up to 48 channels, whereas MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode
- More sampling frequencies than MP3 (from 8 to 96 kHz in contrast to 16 to 48 kHz)
- Arbitrary bit-rates and variable frame length.
- Higher coding efficiency for stationary signals (AAC uses a block size of either 1024 or 960 samples, that allows more efficient coding than for MP3 576 sample blocks)
- Higher efficiency and simpler filter bank (AAC uses a pure MDCT)
- Higher coding accuracy for transient signals (AAC may switch to a block size of 128 or 120 samples, that allows more accurate coding than MP3 192 sample blocks)
- Better coding of frequencies above 16 kHz

- AAC introduces additional tools in order to increase the compression efficiency: TNS, Backwards Prediction, PNS etc.
- More flexible joint stereo (different methods are used depending on frequency ranges)

AAC encoder separates the audio signal in different frequency bands, and then uses psychoacoustic modelling to identify those bands whose complexity can be reduced without audible distortions. This process is more or less aggressive, depending on the constraint imposed by the output bitrate. As a consequence, the audio signal will be more or less deteriorated.

As has been said previously, different implementations of AAC codec result in different audio quality of the encoded audio signal. In order to identify the codec with the highest quality and the corresponding optimal bit-rate, we have performed objective and subjective audio quality evaluation, as described in the sequel.

2.2 Objective audio quality evaluation

The main goal of EAQUAL tool (based on the ITU-R recommendation BS.1387) is to provide an objective quality measure for encoded/decoded audio files. It is not aimed to replace subjective listening tests but to act as a useful tool to support subjective tests and categorize different coding algorithms. The more input files are taken for the analysis, the better the results of EAQUAL will fit the real audio quality. In order to use EAQUAL, one has to provide the reference file, which is the original PCM data (16 bit, 48 kHz format), and the test file, which is the encoded and decoded audio file and has the same audio format (PCM).

EAQUAL uses the psychoacoustic model to compute the signal that would be present at the basilar membrane (excitation pattern), by using nonlinear sum of masks. Afterwards, the cognitive model as well as the combination of different algorithms is used to estimate the impact of noise and distortion due to lossy audio compression. Each of these algorithms provides the so called MOV (Model Output Variable), that are explained as follows:

- Bandwidth: estimate of the bandwidths of both, the reference and the test signal
- NMR (Noise-To-Mask-Ratio): ratio between the noise (introduced by the codec) and the allowed masking threshold
- RDF (Relative Disturbed Frames): based on NMR. If the NMR of any frequency band is higher 1.5dB the frame is assumed to be disturbed. The number of disturbed frames is then divided by total number of frames.
- MFPD (Maximum Filtered Probability of Detection): a detection probability of hearing the noise/artefacts is calculated similarly to the NMR. This probability is smoothed over time by a low pass filter of the first order. After that, the maximum of the resulting value and the MFPD of the preceding frame is chosen.
- ADB (Average Distorted Block): Similarly to RDF, the number of distorted blocks is calculated based on the detection probability (if detection probability > 0.5). Then a measure of how audible the distortion is, is divided by this number.
- EHS (Harmonic Structure of Error): a measure how tonal the noise signal is. The calculation is based on the autocorrelation of the error spectrum.
- Modulation difference: measurement of differences between the envelopes of reference and test signal based on a simple loudness measure and low pass filtering.
- Noise loudness: a measure of the loudness of the noise signal.

The average of these MOVs over time and channels are used as the inputs to a neural network [3]. The output of the network is the DI (Distortion Index) that is mapped through a nonlinear (sigmoid) function to the ODG (Objective Difference Grade) that is shown in Table 1.

Table 1: Objective Difference Grade – ODG.

Impairment description	***ODG***
Imperceptible	0
Perceptible, but not annoying	-1
Slightly annoying	-2
Annoying	-3
Very annoying	-4

2.3 Subjective audio quality evaluation

Subjective audio quality evaluation is performed by using the double blind A/B/X testing methodology [6]-[10]. The testing procedure itself assumes that two different audio recordings are played with random choice as A and B, and one of them is randomly chosen to be X. The listener has to select which recording of A or B is identical to the X.

For the purpose of the A/B/X listening tests, a testing computer program for subjective evaluation was developed in MATLAB [11]. The graphic user interface of the program is shown in Figure 1. There are three pushbuttons — A, B, and X — which play the corresponding audio signal when pushed. It is possible to toggle between the signals during playback and to set the overlap in seconds (the number of seconds the previous playback overlaps with the new one). Markers in the figure that represent the signal in the time domain can be moved easily at any point of the signal and the marked part of the signal between the markers is played. Below the position slider, the subject can monitor the information about current and total time of audio recording. The subject has to decide whether there is a difference between A and B or not. If there is any audible difference, the subject has to decide whether X corresponds to A or B, and finally which signal sounds better (with three radio buttons "I prefer A", "I prefer B" or "I am not sure"). By setting the slider position (at the bottom right corner) the subject provides the subjective difference grade between A and B with the following choices: imperceptible and clearly perceptible. After all decisions have been made, pushing the "continue" button loads the next pair of recordings. Two signals of the same audio material (encoded by two different AAC codecs) are randomly loaded into A or B, so that any subjective influence on the next material is avoided. Whether signal X represents signal A or signal B is randomly determined by the program. Once the testing is completed, the results are saved in a text file named <yyyymmddhhmm.sts>. In such a way, subject's anonymity is preserved, which is of great importance especially when the subjects are experts in the field of audio recording, as it was the case here.

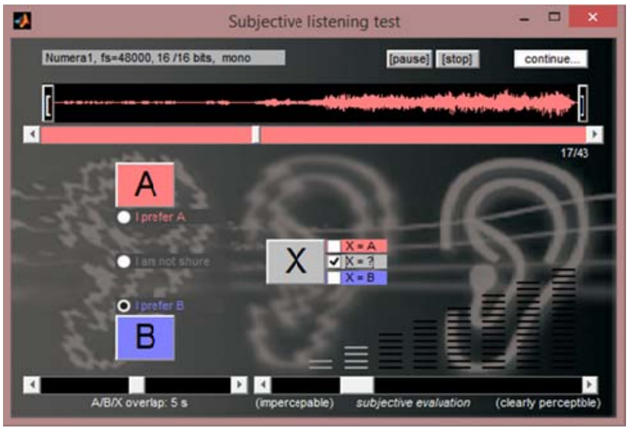

Figure 1. MATLAB application for the subjective audio quality testing.

2.4 Hardware architecture

The main requirement of developed audio device is to provide high audio quality with low energy consumption. Low hardware computing capabilities and battery capacity were identified as the main constraints during its development.

Having in mind the aforementioned constraints, the first choice to make was whether to use hardware or software encoder. Since the first choice requires dedicated chip for audio encoding/decoding, it was immediately discarded as less flexible and more costly, and in some cases less energy efficient. As for the codec choice, AAC codec was preferred because of its advantages in terms of audio quality when compared to MP3.

For the development platform, we have chosen ARM, because of its low price and excellent developer support due to its omnipresence. In addition, the ARM as a RISC architecture is more efficient that similar CISC architectures. ARM Cortex architecture is the most recent realization of ARM microprocessors and it is available in different series: A, M, and R. Series A represent high performance microprocessors, aimed for more demanding operating systems such as Linux, whereas M series have more modest capabilities, and it is aimed for applications without OS.

AAC encoding is a complex process that requires signal processing in time and frequency domain. Both A and M microprocessor series are equipped with a special hardware unit dedicated for signal processing tasks with minimal usage of central processor. Cortex A series possess NEON, SIMD (Single Instruction Multiple Data) unit whereas Cortex M has FPU (Floating Point Unit).

Beside basic features, AAC audio format supports additional tools aimed to improve the audio quality, like TNS (Temporal Noise Shaping), PNS (Perceptual Noise Substitution), MS (Mid-side Stereo), and IS (Intensity Stereo). Since the development of AAC codec is long process, it was decided to evaluate the quality of software solutions already existing on the market, that are optimized for ARM architecture, use hardware unit for signal processing (NEON, FPU), and support specific tools for better encoding such as TNS, PNS, etc.

By analyzing the customer requirements, it has been decided to use AAC-LC (Low Complexity) audio codec. This codec type fits very well the requirement for the output bitrate, and at the same time it is less computationally intensive. Additional features (TNS, PNS) require more processing time, and they were shown to be unnecessary. For instance, PNS requires 10% more processing time, which was not acceptable from the device design aspect.

By surveying the current market, four different AAC codec vendors that fulfil the project requirements have been found. Their software solutions have been evaluated on the hardware platforms based on ARM microprocessors. It must be noted that the solution of Vendor 4 is based on ARM9E architecture that supports a number of DSP instructions and accelerates their execution.

The results representing the complexity of the given solutions are shown in Figure 2. It must be noted that the data shown in Figure 2 correspond to the case of audio recordings with fs=48kHz sampling frequency and the output bitrate of 128Kbps and TNS turned on for the vendors that support it. Only the computational requirements for Vendor 1 are given for ARM platform and audio recording with sampling frequency of fs=96kHz, whose processing requires more computational resources.

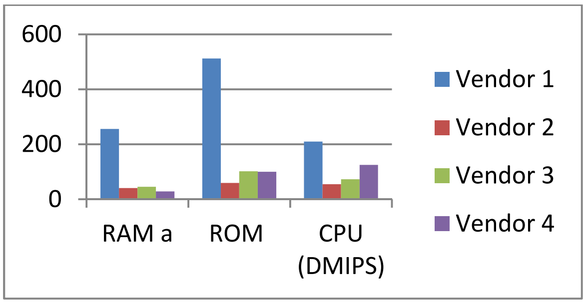

Figure 2. Computational resources consumption.

3 Results

3.1 Objective audio quality evaluation

Test set consisted of 14 audio recordings (mono, fs=48 kHz, 16 bit), encoded and decoded by AAC codecs from four different vendors (See Table 2). Original audio files were encoded to AAC format with the bitrates of 96, 128, 160, and 256 kbps. These files were decoded to .wav format and then compared to the original ones. The result of this comparison is the objective difference grade (ODG). The results averaged over the whole set of 14 recordings are presented in Figure 3.

Table 2. Audio recordings for objective tests

Number	*Recording*
1	Flute Solo
2	Piano, flugelhorn, violin, cello
3	Benjamin Britten - Simple Symphony
4	Mezzo soprano & piano
5	Accordion and organ in a church
6	Flute, cembalo, baroque cello
7	Flute, cembalo, baroque cello
8	Flute, cembalo, baroque cello
9	Piano
10	Piano
11	Acoustic guitar, accordion and vocal
12	Percussion set, Drums duet
13	Percussion set, Drums duet
14	Percussion set, Drums duet

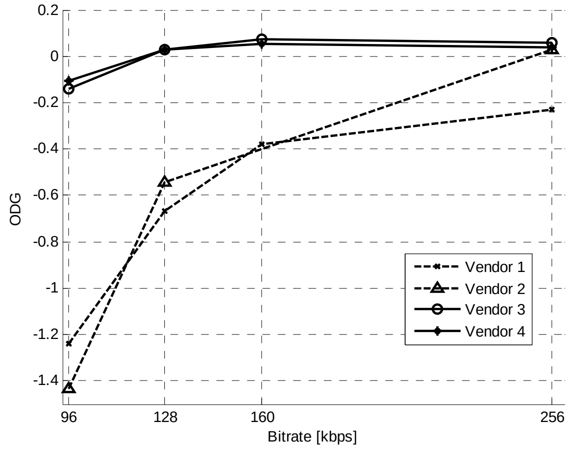

Figure 3. Objective difference grade as a function of bit-rate for 4 different AAC codec vendors.

As can be seen in Figure 3, AAC codecs for Vendors 3 and 4 perform significantly better than codecs for Vendors 1 and 2. For higher bitrates (256 Kbps), all codecs but Vendor 1 codec provide excellent audio quality. Having in mind that the compressed audio recordings will be transferred over wireless connection, the quality of audio recording with lower bitrates was more important. Consequently, codecs provided by Vendors 3 and 4 have been selected for subjective audio quality evaluation.

It must be noted that the Vendor 4 provides two types of AAC codecs. The first one cuts the audio spectrum above 17 kHz before encoding the signal. The second one compresses the original signal (having spectrum up to $fs/2$). The comparison of these two codec variants is given in Figure 4.

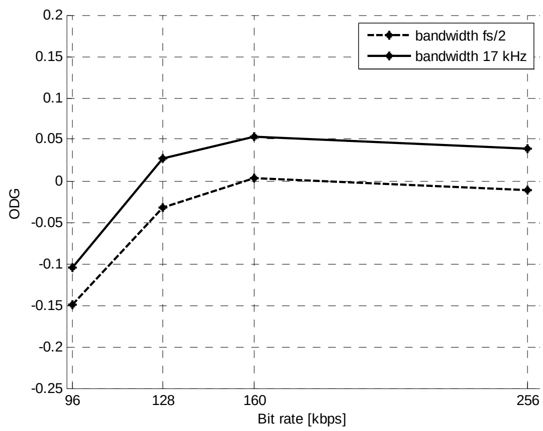

Figure 4. Objective differene grade for two codec variants for Vendor 4.

As can be seen from Figure 4, codec variant that performs low pass filtering provides slightly better quality. This is mainly due to the fact that better bit resolution is achieved at lower harmonics, and the effect of shorter code words does not have as much as impact. Since most of the audio spectrum lies in frequencies lower than 17 kHz, this codec has better audio quality. Beside the fact that human auditory system barely perceives frequencies above 17 kHz, the choice of this algorithm is supported by its lower computational complexity and less processing time.

In Figures 5 and 6, we present ODG as a function of track number for bitrates 96 kbps and 256 kbps respectively. As can be seen, for some tracks codec from Vendor 3 outperforms the codec from Vendor 4 and vice versa. Also, there exist tracks (e.g. Track 6, 13 and 14) that seem to be harder to compress, which results in lower quality of decoded audio. In contrast, Track 12 seems to be quite easy to compress, resulting in similar ODG for all 4 codecs.

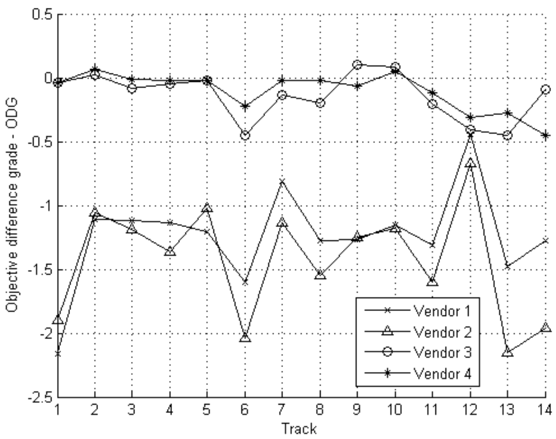

Figure 5. ODG vs. Track number for different codec vendors (Bit rate=96 Kbps)

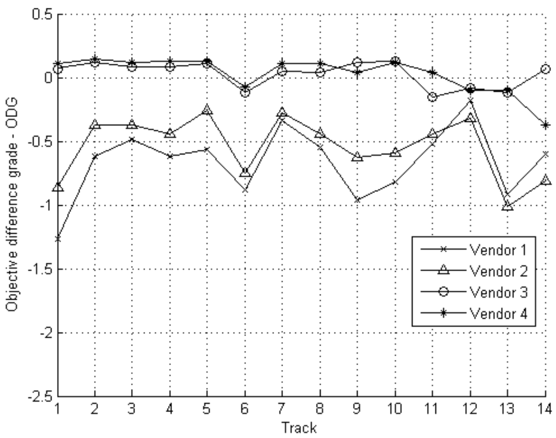

Figure 6. ODG vs. Track number for different codec vendors (Bit rate=256 Kbps)

3.2 Subjective audio quality evaluation

In order to perform subjective testing of audio quality, ten audio recording were chosen. Among them, five were raw tracks and five were produced, as it is shown in Table 3.

Table 3: Audio recordings for subjective tests

Number	*Recording*	*Recording produced*
1	Accordion and organ in a church	YES
2	Piano	YES
3	Acoustic guitar, accordion and vocal	YES
4	Set of percussions	YES
5	Drum set	YES
6	Bass guitar	NO
7	Guitar 1	NO
8	Guitar 2	NO
9	Snare close	NO
10	Female vocal	NO

All subjective tests were performed by using high quality reproduction hardware:

- Laptop PC, ASUS Zenbook U500VZ serial number: D7M0CY18988672A 24M,
- USB D/A converter with built-in headphone amplifier, audiolab M-DAC serial number: AH001841BFC1076,
- Headphones MB Quart QUART PHONE 400 serial number: 14444.

Subjective tests involved six audio professionals, two acousticians and three young listeners, all aged from twenties to sixties. The subjects performed test individually. The single test duration was approximately 20 min.

Ten previously chosen audio recordings from the Table 3 were encoded and decoded with AAC codecs from Vendors 3 and 4 with the corresponding bit-rates. Two different tests were performed: testKK in which two codecs with 96kbps were compared and the one was chosen as a better, and test96 in which the better codec (96kbps bitrate) was compared to the original raw recording. The details of the tests together with the results are summarized in Table 4.

According to the results of the subjective evaluation of the two selected codecs with seven subjects, there is no significant difference between the two selected codec implementations.

Table 4: Subjective test results.

Test	*Sample1*	*Sample 2*	*Number of subjects*	*Count of correct choice for X*	*Count of undefined X*	*Grade: „1 is better“*	*Grade: „2 is better“*	*Grade: „no difference“*
TestKK	Codec 3 48 kHz, 16 bits, 96 kbps	Codec 4 48 kHz, 16 bits, 96 kbps	7	27/70 38.6%	31/70 44.3%	24	34	12
Test96	Codec 4 48 kHz, 16 bits, 96 kbps	Original 48 kHz, 16 bits, 768 kbps	6	13/60 21.7%	27/60 45%	24	17	19

In conversation with the subjects after each test, they claimed that the difference is quite small, and that, in some cases, they prefer low frequency response of codec 3 (bass guitar and drum samples) and, in some cases, they prefer mid/high frequency response of codec 4 (voice, acoustic guitar, piano).

The overall impression was that the quality of the sound can be evaluated as very good, and the difference may be noticed only in direct comparison of the samples.

According to the results of the subjective evaluation of bit rate effect on perceived quality (codec 3, 96 kbps), there is a certain difference between the selected codec implementation and the original, but it cannot be considered as significant. Summary impression of the quality of the chosen codec versus original samples is that the difference may be heard only in direct comparison with the original.

Overall, the final conclusion after the test and discussion with professionals is that both codecs under test are appropriate for the design, even at the lowest bit rate.

4 Conclusion

In this paper, we present an evaluation of different AAC audio codec implementations for ARM architectures. By performing objective audio quality tests using EAQUAL tool, we have selected two best performing codec realizations. Next we have performed subjective audio testing in order to identify the optimal codec. Results of the subjective testing have shown that the difference may be heard only in direct comparison with the original.

References:

[1] K. C. Pohlmann Principles of Digital Audio, 5th ed. McGraw-Hill, New York, 2005.

[2] Gilbert A. Soulodre, Theodore Grusec, Michel Lavoie, Louis Thibault, Subjective Evaluation of State-of-the-Art 2-Channel Audio Codecs, Proc. of the 104th Audio Engineering Society Convention, May 1998

[3] A. Lerch, Zplane Development, Eaqual-Evaluate Audio Quality, Version 0.1.3alpha

[4] ITU-R Recommendation BS.1387-1 (11/01) Method for objective measurements of perceived audio quality http://www.itu.int/rec/R-REC-BS.1387-1-200111-I/en

[5] Karlheinz Brandenburg, MP3 and AAC explained, AES 17th International Conference on High Quality Audio Coding, Aug. 1999

[6] D. L. Clark, High resolution subjective testing using a double-blind comparator, J. Audio Eng. Soc. Vol. 30, No. 5, 1982

[7] D. L. Clark, Ten years of A/B/X Testing, Proc. of the 91st Audio Engineering Society Convention, May 1991

[8] G. A. Soulodre, M. Lavoie, Subjective evaluation of large and small impairments in audio codecs, Proc. of the AES 17th. International Conference on High-Quality Audio Coding, Aug.1999

[9] J. Hynninen, N. Zacharov, GuineaPig - A Generic Subjective Test System for Multichannel Audio, Proc. of the 106th Audio Engineering Society Convention, May 1999

[10] ITU-R Recommendation BS.1116-1(85/10), Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

[11] I. M. Salom, D. Z. Todorović, Lj. D. Milić, The Influence of Impulse Response Length and Transition Bandwidth of Magnitude Complementary Crossovers on Perceived Sound Quality”, Journal of the Audio Engineering Society, November 2007, Vol. 56, No. 11