

Comparison of audio codecs using PEAQ algorithm

Marija Šalovarda, Ivan Bolkovac, Hrvoje Domitrović
Faculty of E.E. and Computing, Dept. of Electroacoustics, Zagreb, Unska 3

Summary

Until recently, the only widely accepted assessment procedures for audio or speech codecs were listening tests, due to the lack of international standards for measuring the perceived audio quality. Consequently, the idea of substituting the subjective tests by objective, computer based methods has been an ongoing focus of research and development. With new measurement algorithm PEAQ we have measured a characteristics of four different codecs and have shown the results in this paper.

1. Introduction

Compression has become state-of-the-art technology in modern audio communications, both in wide-band and voice-band, thus allowing for a great number of components such as mobile phones, radio and TV satellite networks, Internet audio, digital audio broadcasting below 30 MHz (DRM – Digital Radio Mondiale) and over 30 MHz (DAB - Digital Audio Broadcasting), DVD (Digital Versatile Disc), VoIP (Voice Over Internet Protocol), and many more. Lowering data rates to a minimum is contradictory to clarity and fidelity of sound. In time of “all digital technology”, sound quality and the intelligibility of speech have become important again.

However the “old fashioned” methods of measuring, i.e. objective grading the actual audio communication system are not satisfactory for those systems. The traditional methods deal with S/N and distortion (linear and nonlinear) measurements. Today audio encoders that uses compression, or better to say – reduction, uses psychoacoustic models of human hearing as the base for bit-rate reduction, so it is necessary to simulate the subjective evaluation of human subjects if we want to judge these systems. The analysis is based on the most recent perceptual techniques, such as PEAQ (Perceptual Evaluation of Audio Quality), PESQ (Perceptual Evaluation of Speech Quality) and PSQM (Perceptual Speech Quality Measurement). Since the psychoacoustic models are developed upon investigations with the real signal, the contemporary measurements have to use the same natural stimulus for measurement: human voice and music program material.

Employment of such stimulus makes it possible to monitor the quality during normal operation of system under test. As a consequence of this approach that measures the perceived audio quality instead of signal characteristics, it is possible to gain an objective metrics, which truly characterizes the quality of service (“QoS”) of a system or a network.

Due to the lack of international standards and recognized measurement procedures, the only widely accepted assessment procedures for audio or speech codecs were listening tests. The first methods for testing telephone band speech signals were standardized within ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) Recommendation P.800 in 1993. In 1994. ITU has recommended the ITU-R BS.1116 (ITU-R - ITU Radiocommunication Sector) - a test procedure to assess wide band audio codecs on the basis of subjective tests.

The analysis of the results from a subjective listening test is based on the Subjective Difference Grade (SDG) defined as:

$$SDG = Grade_{\text{Signal under test}} - Grade_{\text{Reference signal}}$$

The values of the SDG range from 0 to -4, where 0 corresponds to an “imperceptible impairment” and -4 to a “very annoying impairment”.

Both described methods use results from listening tests that are time-consuming, expensive and impractical for everyday use. Consequently, the idea of substituting the subjective listening tests with objective, computer-based methods has become the main goal. After few years of research, in 1996 ITU finalized the recommendation P.861. This recommendation defines the method for the objective analysis of speech codecs. Since it is correlated up to 98% with the scores of subjective listening tests, it uses the PSQM algorithm for cognitive perceptual model. For specific applications, such as VoIP, the PSQM algorithm is not satisfactory, so in 2001. ITU finalized another refined method through recommendation P.862. It uses the PESQ algorithm for cognitive perceptual model. For wide band audio codecs ITU-R recommended the PEAQ algorithm implemented in recommendation ITU-R BS.1387.

2. Measurement description

The underlying concepts of the proposed algorithms for perceptual measurement techniques are all quite similar. The common structure of these algorithms is depicted in Figure 1

The process of human perception is modelled by employing a difference – measurement-technique that compares both, a reference signal (i.e. the "input" signal to a codec) and a test signal (i.e. the "output" signal of the codec).

First, the algorithms process an ear model for the reference and the test signal, in order to calculate an estimate for the audible signal components. The result can be imagined as the "internal representation" inside the human auditory system. Comparing the internal representations of the reference, and the test signal leads to an estimate of the *audible difference*. To derive an overall quality figure, this information, which is a function of time, must be processed accordingly, like the human brain of a subject would do in a listening test. The respective part of processing within an algorithm is referred to as *cognitive modelling*. A total quality figure will be derived as the final result, which can be compared to a MOS ("Mean Opinion Score") resulting from a listening test.

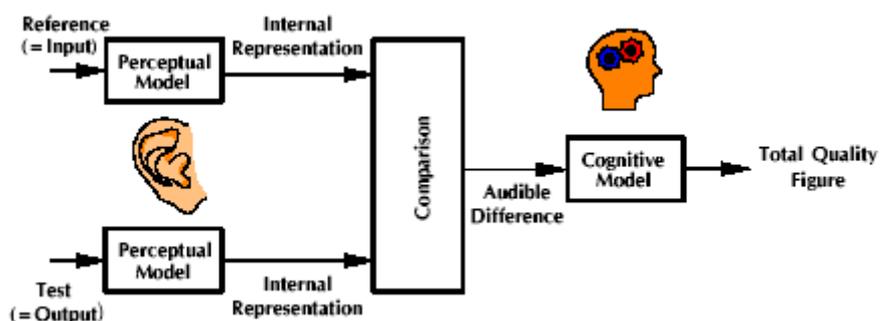


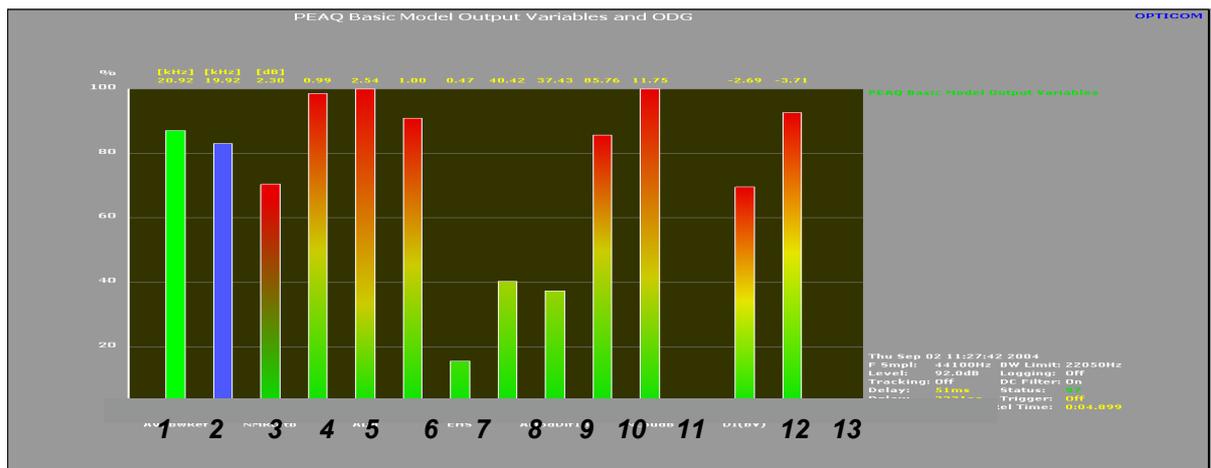
Figure 1-The underlying concept for perceptual measurement

2.1. Model Output Variables

The process of comparison, based upon mentioned cognitive models, is divided into several phases. Each phase gives us as a result one or more Model Output Variables (MOV), i.e. descriptors for various cognitive processes. The final quality figure takes into account all MOVs and is represented as single number Objective Difference Grade (ODG).

Description of the Model Output Variables

OPERA name	Model Output Variable (MOV)	Description
WModDif1B	WinModDiff	Windowed averaged difference in modulation (envelopes) between Reference Signal and Signal Under Test
AmodDif1B	AvgModDiff1	Averaged modulation difference
AmodDif2B	AvgModDiff2	Averaged modulation difference with emphasis on introduced modulations and modulation changes where the reference contains little or no modulations
NLoudB	RmsNoiseLoud	Rms value of the averaged noise loudness with emphasis on introduced components
AvgBwRef	BandwidthRef	Bandwidth of the Reference Signal
AvgBwTst	BandwidthTest	Bandwidth of the output signal of the device under test
NMRtotB	TotNMR	logarithm of the averaged Total Noise to Mask Ratio
RDF	RelDistFrames	Relative fraction of frames for which at least one frequency band contains a significant noise component
MFPD	MFPD	Maximum of the Probability of Detection after low pass filtering
ADB	ADB	Average Distorted Block (=Frame), taken as the logarithm of the ratio of the total distortion to the total number of severely distorted frames
EHS	EHS	Harmonic structure of the error over time



- 1-AvgBwRef*
- 4-ADB*
- 7-RDF*
- 10-AModDif2B*
- 2-AvgBwTst*
- 5-MFPD*
- 8-WModDif1B*
- 11-NLoudB*
- 3-NMRtotB*
- 6-EHS*
- 9-AModDif1B*
- 12-DI (Distortion Index)*
- 13-ODG*

Figure 2 - PEAQ Basic Model Output Variables (MOVs) and the ODG

3. Results

In this paper we show the results of tests made on various codecs using the PEAQ measurement algorithm according to ITU-R BS.1387. The codecs used are standard MP2 and MP3 (or MPEG 1 layer 2 and layer 3, according to ISO/IEC 11172/3, 1992. where MPEG stands for Moving Picture Experts Group and ISO/IEC for International Standards Organization/International Electrotechnical Commission), AAC (Advanced Audio Coding or MPEG 2 AAC, according to ISO/IEC 13818/3, 1994.) and OGG Vorbis (free codec from Xiph.org, different from MPEG 1 and 2 standards). They all use psychoacoustic methods for lossy compression (some data is discarded as irrelevant according to their algorithm and cannot be restored unlike lossless compression where no data is lost) of the audio data - the differences are in their algorithms and their complexity so they do not act the same which is visible from the results of the measurements.

We made the measurements using the computer measuring system Opera from Opticom. The audio clip used for encoding and testing was ripped from audio CD in WAV format (16 bit, 44.1 kHz), it was 5 seconds in duration, and was the same for all four codecs.

The measurement was done in the following way: first we encoded referential audio clip on all codecs and on all most common bitrates. Then we decoded all of the resulting compressed clips and made tests in Opera comparing them with the referential uncompressed clip. In this phase the results are showed only for the 100% compatible encoders and decoders, i.e. the devices from the same manufacturer. Please note that all measurements were done in stereo so the bit rates are showed accordingly (i.e. 128 kbps refers to two [left and right] 64 kbps encoded audio channels).

For the brief overall comparison the final ODG value for different bit rates are shown in Fig.3. For deeper analysis we used the MOVs obtained during the measurements.

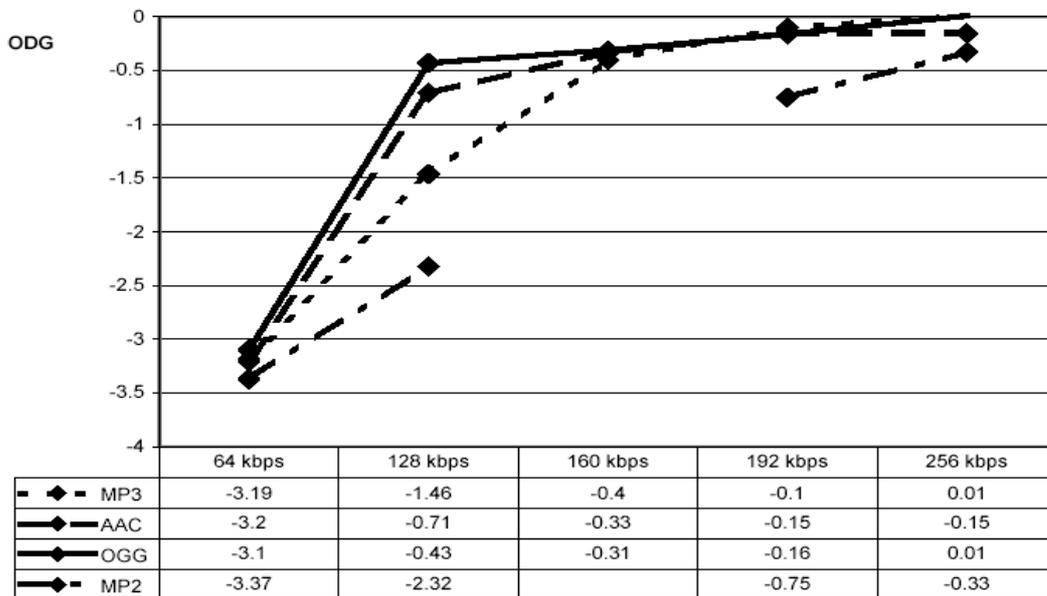


Figure 3 – ODG values for 4 different codecs

The screen shots in next few figures (Fig. 4,5,6,7) show the MOVs measured on four codecs on 64 kbps bitrate.

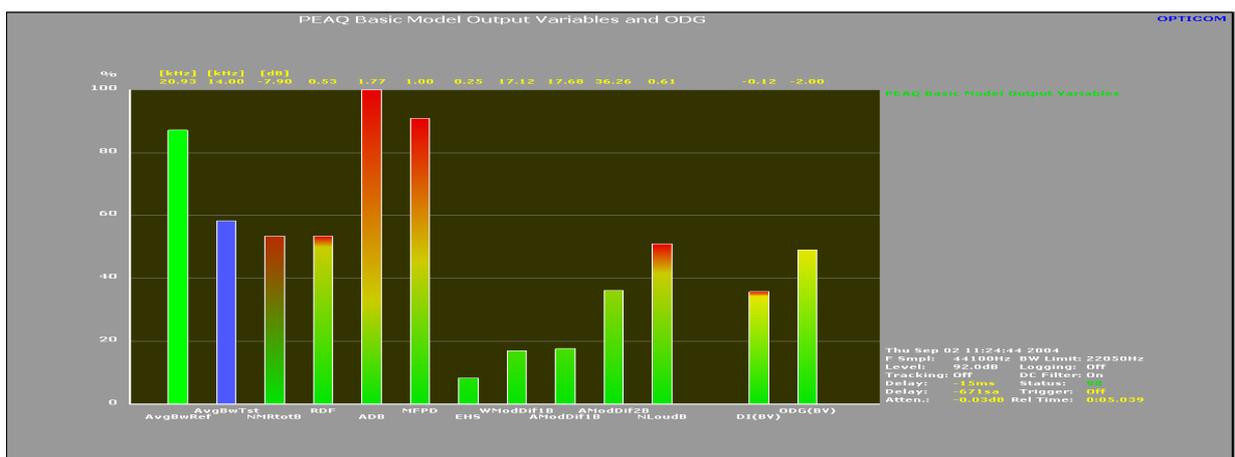


Figure 4 - MP2 MOVs

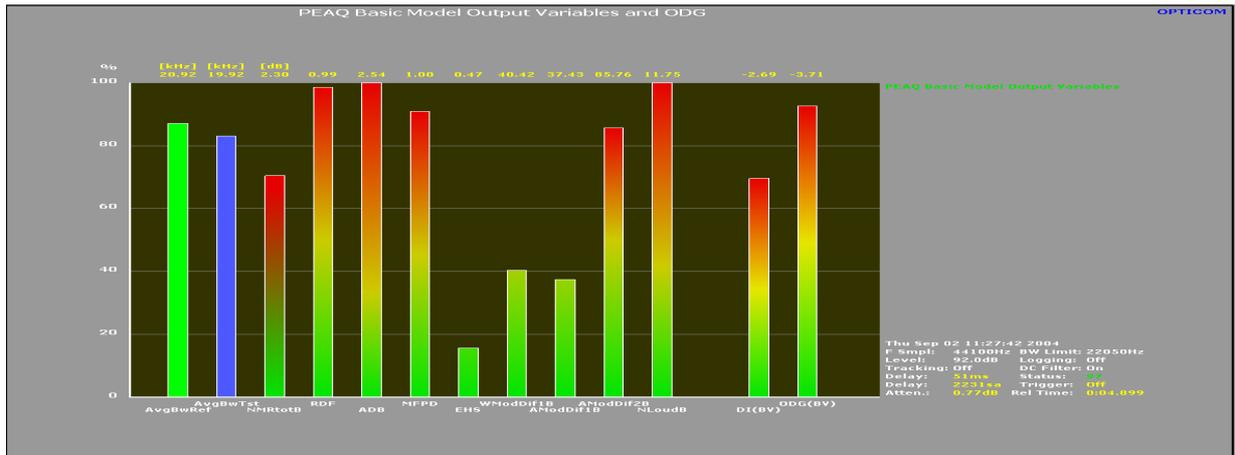


Figure 5 - MP3 MOVs

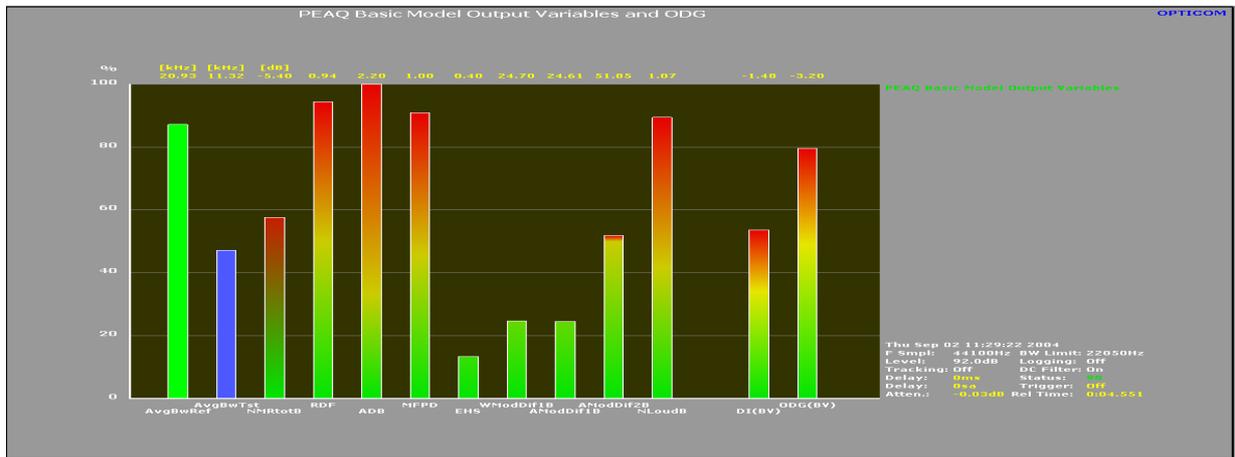


Figure 6 - AAC MOVs

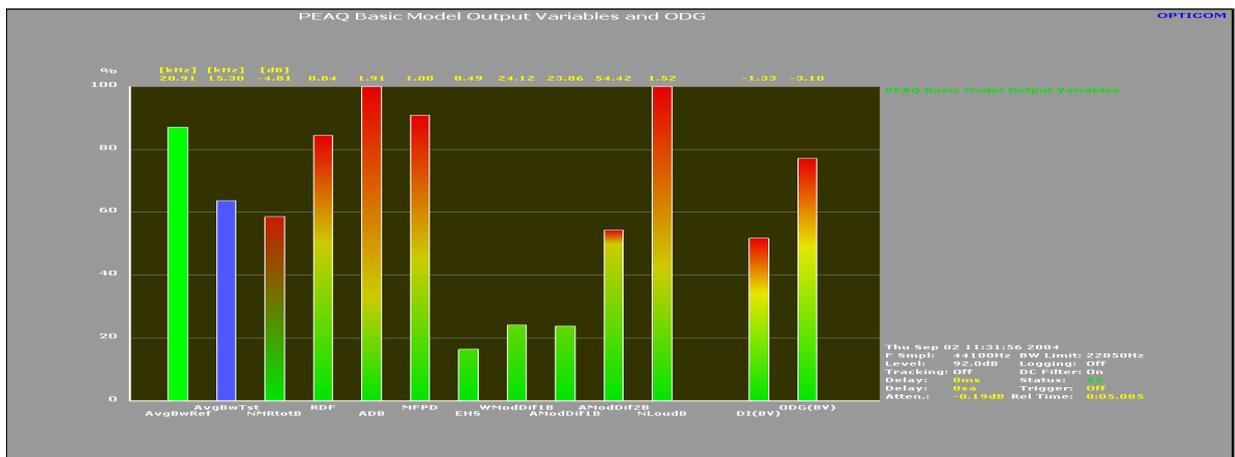


Figure 7 - OOG MOVs

4. Conclusion

It is visible that all codecs act similarly at higher (>160 kbps) bitrates - the differences are minimal except for the MP2. On the lower bitrates (<160 kbps) on the other hand, we can see different behavior of all four codecs, especially in the most interesting 128 kbps. The best is OGG Vorbis and the AAC is very near.

From these results we can conclude that it is very important to pick the right codec at lower bitrates while it is not so important on higher bitrates in the terms of audio quality. However, due to the popularity of some codecs, i.e. MP3, the choice is often not only based on quality and that is why it is very hard to recommend any in particular.

Acknowledgment

The authors would like to thank to the company Dekod telekom d.o.o. from Zagreb for providing us the measurement system Opticom OPERA 3.5, as well as the excellent support in the beginning tests.

References

- [1] T.THIEDE AND AL., *PEAQ The ITU Standard for Objective Measurement of Perceived Audio Quality*, J.Audio Eng.Soc., Vol. 48, No 1/2, 2000., pp. 3-29
- [2] W.C.TREURNIET, G.A.SOULODRE, *Evaluation of the ITU-R Objective Audio Quality Measurement Method*, J.Audio Eng.Soc., Vol. 48, No 3, 2000., pp. 164-173
- [3] xxx, OPERA Your Digital Ear ver. 3.5, User Manual, Opticom, 2002.