Estimating Perceptual Audio System Quality Using PEAQ Algorithm

Marija Šalovarda, Ivan Bolkovac, Hrvoje Domitrovic
Faculty of Electrical Engineering and Computing, Unska 3, HR – Zagreb, Croatia,
E-mail: *marija.salovarda@dekod.hr, ibalki@hotmail.com, hrvoje.domitrovic@fer.hr*

Abstract

The idea of objective, computer based methods for testing audio quality is introduced in this work. Including some basic of perceptual coding, international standards and description of new measurement algorithm. With described 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.

Today audio encoders are using psychoacoustic models of human hearing as the base for bit-rate reduction. 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 that cannot be applied on modern codecs because they are non-linear and non-stationary. For objective evaluation it is necessary to simulate the subjective evaluation of human subjects if we want to judge these systems.

Most common objective analysis is based on the 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.

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. [1]

2. PERCEPTUAL CODING

The main question in perceptual coding is: how much noise can be introduced into the signal before it becomes audible? Answers to this question are derived from psychoacoustic. Psychoacoustic describes the relationship between acoustic events and resulting perceived sounds. [2]

The most important keyword is masking. It describes the effect by which a fainter, but distinctly audible, signal becomes inaudible when correspondingly louder signal occurs simultaneously, or within a very short time.

Masking depends on the spectral composition of both the masker and masked signal, and on their variations with time. The basic structure of perceptual coding is shown is in figure 1. A filter bank is used to decompose the input signal into subsampled spectral components. Together with a corresponding filter bank in the decoder it forms an analysis and synthesis system. Using either the time-domain input signal or the output of the analysis filter bank, an estimate of the actual masked threshold is computed using rules known from psychoacoustic. This is perceptual model of the encoding system. The spectral components are requantized and coded with the aim of reducing the resolution of the coding while keeping the quantization noise below the masked threshold. Finally a bit-stream formatter is used to assemble the bit stream, which typically consists of quantized and coded spectral coefficients and some side information. [3]

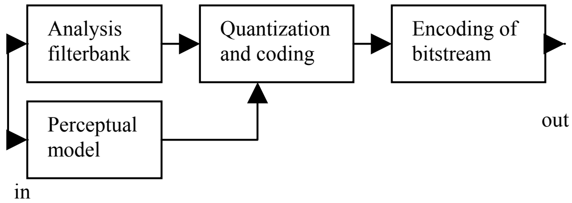

Figure 1. Basic block diagram of perceptual coding system

3. INTERNACIONAL STANDARDS

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.

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. [4][5]

4. OBJECTIVE AUDIO QUALITY MEASUREMENTS

Since subjective listening tests are time-consuming, expensive and impractical for everyday use it was beneficial to substitute the subjective listening tests with objective, computer-based methods.

The analysis of the results from a subjective listening test is based on the Subjective Difference Grade (*SDG*) [4][5] defined as:

$$SDG = Grade\text{Signal under test} - Grade\text{Reference signal} \tag{1}$$

Table 1 SDG description

Impairment	Grade	SDG
imperceptible	5,00	0,00
perceptible, but not annoying	4,00	-1,00
slightly annoying	3,00	-2,00
annoying	2,00	-3,00
very annoying	1,00	-4,00

The Objective Difference Grade *ODG* is the output variable from objective measurement method and corresponds to the *SDG* in the subjective domain.

4.1 PEAQ

PEAQ measurement method models fundamental properties of the auditory system. Several intermediate stages model physiological and psychoacoustic effects.

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*. [4][5][6]

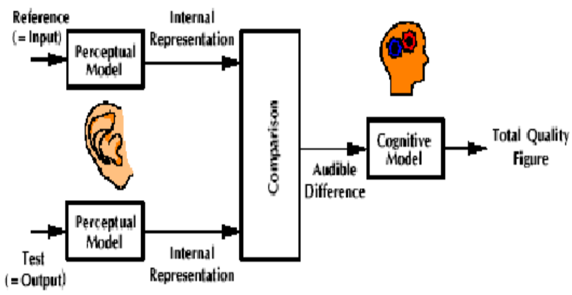

Figure 2. The underlying concept for perceptual measurement

4.1.1. Model output variables (*MOV)*

The PEAQ measurement model produces a number of variables based on comparison between a reference signal and the same signal processed by particular device, in our case codec. These variables are called Model Output Variables *MOV*. The final stage of the measurement model combines the *MOV* values to produce a single output value that directly corresponds to an expected result from a subjective quality assessment – *ODG*. The function that maps the model variables to the *ODG* was calibrated using audio items and subjective ratings from a number of listening tests that used the ITU-R BS.1116 methodology. [6]

Table 2 MOV description

OPERA name	BS.1387 name
AvgBwRef	Average Bandwidth of the Reference Signal
AvgBwTst	Average Bandwidth of the output signal of the device under test
NMRtotB	Total Noise-to-Mask Ratio
ADB	Average Distorted Block (Frame), taken as the logarithm of the ratio of the total distortion to the total number of severely distorted frames
MFPD	Maximum of the Probability of Detection after low pass filtering
EHS	Harmonic structure of error over time
RDF	Relative Fraction of frames for witch as last one frequency band contains a significant noise component
WModDif1B	Windowed averaged difference in modulation (envelopes) between Reference Signal and Signal under Test
AModDif1B	Averaged modulation difference
AModDif2B	Averaged modulation difference with emphasis on introduced modulations and modulation changes where the reference contains little or no modulations
NLoudB	RMS value of the averaged noise loudness with emphasis on introduced components

5. RESULTS

In this paper the results are shown 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 Lame (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 (perceptual coding) 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 10 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).

The screen shots in figures 4-7 are taken out from Opera system and they show the *MOVs* as they are defined by BS.1387. The results are shown framewise and are averaged since the beginning of the measurement. This diagram contains several bars, each having a different meaning, scaling and unit. The percent scale on the left side of the diagram is for orientation only merely. The first 11 bars represent the *MOVs*, the last two bars are distortion index and the final *ODG*.

Table 3 ODG values and file size for all four codecs on most common bit rates

	codec	ODG	file size
32 kbps	MP2	-3.85	56kB
	MP3	-3.67	56kB
	AAC		
	OGG		
64 kbps	MP2	-3.28	111kB
	MP3	-3.46	112kB
	AAC	-3.36	111kB
	OGG	-3.24	113kB
128kbps	MP2	-2.36	222kB
	MP3	-1.08	223kB
	AAC	-1.09	222kB
	OGG	-0.34	221kB
160 kbps	MP2		
	MP3	-0.47	276kB
	AAC	-0.41	278kB
	OGG	-0.20	271kB
192kbps	MP2	-0.59	332kB
	MP3	-0.16	334kB
	AAC	-0.20	333kB
	OGG	-0.09	326kB
256 kbps	MP2	-0.19	443kB
	MP3	-0.01	445kB
	AAC	-0.12	429kB
	OGG	0.02	460kB
320 kbps/ 350kbps -ogg	MP2		
	MP3	0.04	556kB
	AAC	-0.02	535kB
	OGG	0.07	632kB

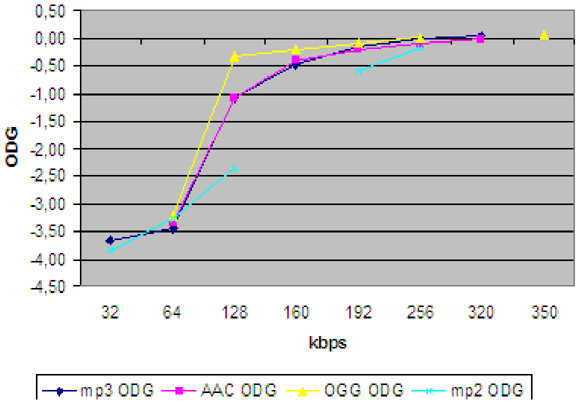

Figure 3. Comparison of the final ODG value of four different codecs on different bit rates

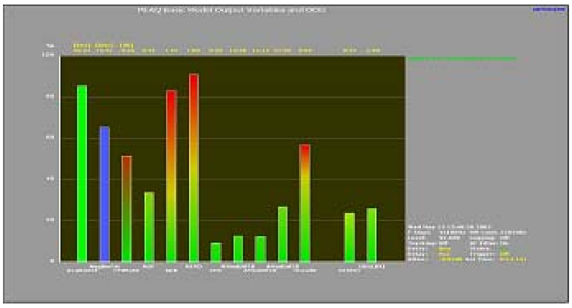

Figure 4 . MP3 Lame MOV values on 128 kbps

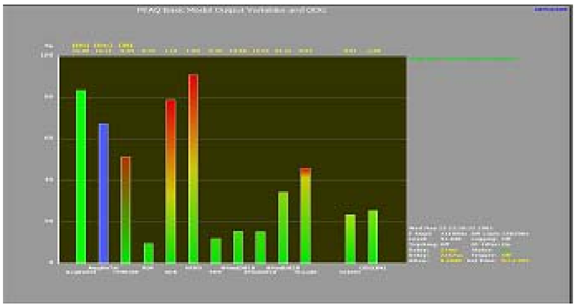

Figure 5. AAC MOV values on 128 kbps

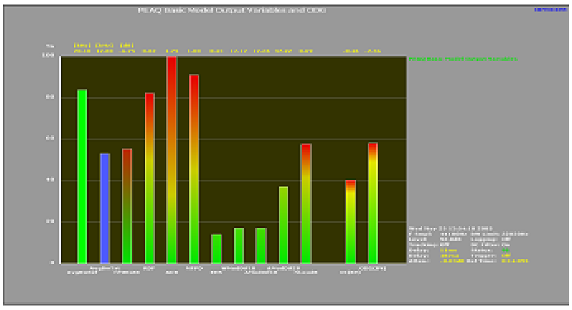

Figure 6. MP2 MOV values on 128 kbps

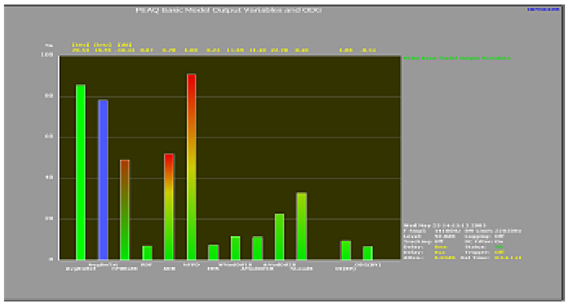

Figure 7. OGG Vorbis MOV values on 128kbps

6. CONCLUSION

It is visible that all codecs act similarly at higher (>160 kbps) bit rates – the differences are minimal except for the MP2.

On the lower bit rates (<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, very similar qualities have AAC and MP3, MP2 has the lowest quality at this bit rate.

From these results we can conclude that it is very important to pick the right codec at lower bit rates while it is not so important on higher bit rates in the terms of audio quality. However, some times it's important not only to se quality of particular codec but also size of encoded file. In table 3 we can see that as higher the quality is (*ODG*), larger is file size.

7. REFERENCES

[1] M.Salovarda, I.Bolkovac, H.Domitrovic "Comparison of audio codecs using PEAQ algorithm", *CUC04*, Zagreb, September 2004.

[2] E.Zwicker and H. Fastl, *Psychoacoustic, Facts and Models,* Springer, Berlin, 1990.

[3] ISO/IEC IS 11172-3, "*Coding of Moving Pictures and Associated Audio of Digital Storage Media at up to 1.5Mbit/s, Part 3:Audio*" (1992).

[4] 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

[5] 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

[6] xxx, OPERA Your Digital Ear ver. 3.5, User Manual, Opticom, 2002.