# PEAQ–An Objective Method To Assess The Perceptual Quality of Audio Compressed Files



Andrei Câmpeanu
Polytechnic University of Timisoara

# PEAQ – AN OBJECTIVE METHOD TO ASSESS THE PERCEPTUAL QUALITY OF AUDIO COMPRESSED FILES

**Dinu Câmpeanu and Andrei Câmpeanu**

*"Politehnica" Univ. Timişoara, Dept. of Communications, Bv. V. Parvan Nr.2, Romania, tel. +40256403309,*
*emails: campeanu@gmail.ro, andrei.campeanu@etc.utt.ro*

Abstract: Up until now, the only way to measure the sound quality of modern audio coding systems at low bit rates has been to implement elaborate listening tests with experienced human test subjects. Consequently, the idea of substituting the subjective tests by objective, computer based methods has been an ongoing focus of research and development. The result was PEAQ, an algorithm (ITU-R Recommendation BS.1387-1), that uses a number of psycho-acoustical tests combined to give a measure of the quality difference between two instances of a signal (a reference and a test signal). The paper describes a MATLAB PEAQ method implementation and the procedure used to assess it. Finally, PEAQ is used to rate the quality of operation for some very known digital audio editors and audio codecs.

Keywords: audio, quality, measurement, evaluation, perceptual, perceived, algorithm, codec.

## 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.

Perceived audio quality is one of the key factors when designing digital audio communication systems. Formal listening tests have been the relevant method for judging audio quality. In 1994 ITU (International Telecommunication Union) has recommended the ITU-R BS.1116 (ITU Radiocommunication Sector) – a test procedure to assess wide band audio codecs on the basis of subjective tests. The analysis of results from a subjective listening test is based on the Subjective Difference Grade(*SDG*) defined as

$$SDG = Grade_{\text{Signal under test}} - Grade_{\text{Reference signal}} \quad (1)$$

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

However, subjective quality assessments are both time consuming and expensive. It was desirable to develop an objective measurement in order to produce an estimate of the audio quality. Traditional methods like signal-to-noise-ratio (SNR) or total-harmonic-distortion (THD) have never really been shown to relate reliably to the perceived audio quality. The problems become even more evident when the methods are applied on modern codecs, which are both non-linear and non-stationary. After through verification, ITU-R recommends an objective measurement method, known as PEAQ (Perceived Evaluation of Audio Quality), to estimate the perceived audio quality of equipment under test, for example a low bit-rate codec. This method is specified as ITU-R BS.1387-1 (ITU, 2001; Thiede *et al*., 2000).

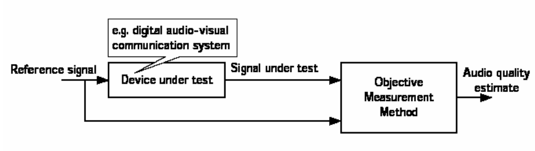



Fig. 1 The basic principle used to carry out objective measurements.

The output variable from the PEAQ objective measurement method is the objective difference grade (ODG) and distortion index (DI). The ODG corresponds to the subjective difference grade (SDG) in the subjective domain. The resolution of the ODG is limited to one decimal. One should however be cautious and not generally expect that a difference between any pair of ODGs of a tenth of a grade is significant. The DI has the same meaning as the ODG. However, DI and ODG can only be compared quantitatively but not qualitatively. As a general rule, the ODG should be used as the quality measure for ODG values greater than approximately –3,6. The ODG correlates very well with subjective assessment in this range. When ODG value is less than –3,6, the DI should be used.

## 2. PEAQ ALGORITHM PRINCIPLES

The basic concept for PEAQ objective measurement method is illustrated in Fig. 1. It consists of two inputs, one for the (unprocessed) reference audio signal and the other for the audio signal under the test. The latter may, for example, be the output signal of digital audio-video communication systems, that is stimulated by the reference signal. This measurement method is applicable to most types of audio signal processing equipment, both digital and analogue. It is, however, expected that many applications will focus on audio codecs.

A high-level representation of PEAQ model is shown in Fig. 2. PEAQ method is based on generally accepted psychoacoustic principles. In general it compares a signal that has been processed in some way with the corresponding time-aligned reference signal. Concurrent frames of the reference and processed signal are each transformed to the outputs of ear models. In a consecutive step, algorithm models the audible distortion present in the signal under test by comparing the outputs of the ear models. The information obtained by these process results into several values, so called MOVs (“Model Output Variables”) and may be useful for detailed analysis of the signal. The final goal instead is to drive a quality measure, consisting of a single number that indicates the audibility of the distortions present in the signal under test. In order to achieve this, some further processing of the MOVs is required which simulates the cognitive part of the human auditory system. Therefore the PEAQ algorithm uses an artificial neural network.

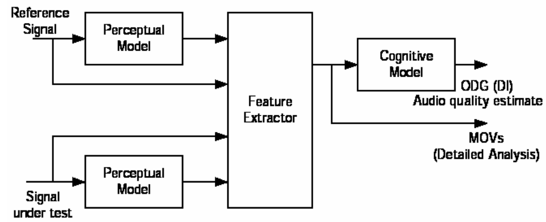



Fig. 2 Representation of PEAQ model

There are two versions of PEAQ, a “Basic” version, featuring a low complexity approach, and an “Advanced” version for higher accuracy at the trade off of higher complexity. The structure of both versions is very similar, and fits exactly into the PEAQ model shown in Fig. 2. The major differences between the “Basic” and the “Advanced” version are hidden in the respective ear models and the set of MOVs used. Next, we will present the most important features of “Basic” version of PEAQ method, the only algorithm used in the paper.

The “Basic” version implements an FFT based ear model, as outlined in Fig. 3. Most features of this model are based on the fundamental psychoacoustic principles. Fig. 3 shows the signal flow from the input signal to the final calculation of the excitation pattern. The processing starts by a transformation of the input signal to the frequency domain. A 2048-point FFT is applied along with subsequent scaling of the spectra, according to the listening level, which has to be input by the user as a parameter. This results in the frequency resolution of approximately 23,4 Hz, and a corresponding temporal resolution of 23,4 ms (at 48 kHz sampling rate).

In the constructive block, the effects of the outer and middle ear are modelled by weighting the spectrum with the appropriate filter functions. Afterwards the spectra are grouped into critical bands, achieving a resolution of 1/4 bark per band. The subsequent adding of “internal noise” is intended to model effects, such as the permanent masking of sounds in our auditory system caused by the streaming of blood and other physiological phenomena. This step is followed by calculation of masking effects. Simultaneous masking is modelled by a frequency and level dependent spreading function. Temporal masking is modelled only partly since the temporal resolution is the same range as the timing of any background masking effects, which therefore cannot

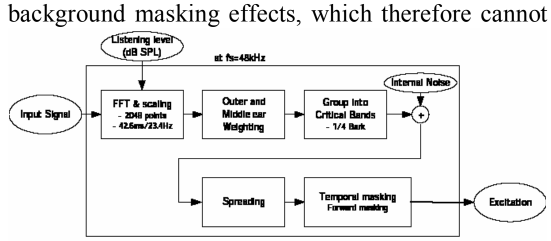



Fig. 3 FFT based ear model used in PEAQ „basic” version

Table 1 Model output variables, PEAQ "Basic" version

| Model Output Variables (MOV) | Description |
|---|---|
| $WinModDiff1_B$ | Windowed averaged difference in modulation (envelopes) between Reference and Test Signals |
| $AvgModDiff1_B$ | Averaged modulation difference |
| $AvgModDiff2_B$ | Averaged modulation difference with emphasis on introduced modulations |
| $RmsNoiseLoud_B$ | Rms value of the averaged noise loudness with emphasis on introduced components |
| $BandwidthRef_B$ | Bandwidth of the Reference Signal |
| $BandwidthTest_B$ | Bandwidth of the Test Signal |
| $RelDistFrames_B$ | Relative fraction of frames for which at least one frequency band contain significant noise |
| Total $NMR_B$ | Logarithm of the averaged Total Noise to Mask Ratio |
| $MFDP_B$ | Maximum of the probability of detection after low pass filtering |
| $ADB_B$ | Average Distorted Block |
| $EHS_B$ | Harmonic structure of the error over time |

be modelled. Nevertheless, experiments have shown that backward masking is very coarsely modelled by side effects of the FFT.

Using the feature extractor, eleven MOVs are extracted from the compensation of the ear model output. Table 1 shows a list of those MOVs and their interpretation. For further information about the MOVs please refer to the annex of the ITU-R recommendation BS.1387.

## 3. A MATLAB IMPLEMENTATION OF PEAQ

There are some papers describing implementations of the PEAQ algorithm in different programming languages. A first implementation was made by Baumgarte and Lerch (2001) for the BASIC version of the algorithm. Kabal (2002) from McGill University implemented parts of PEAQ as C-language and MATLAB versions. We used the MATLAB-code routines submited in Kabal (2002) to develop our own MATLAB implementation of the "basic" version of PEAQ method.

Fig. 4 gives the generic block diagram of MATLAB program which implements PEAQ algorithm. The input signals are WAV-files (Microsoft Riff format), sampled at 48kHz, 16-bit PCM.

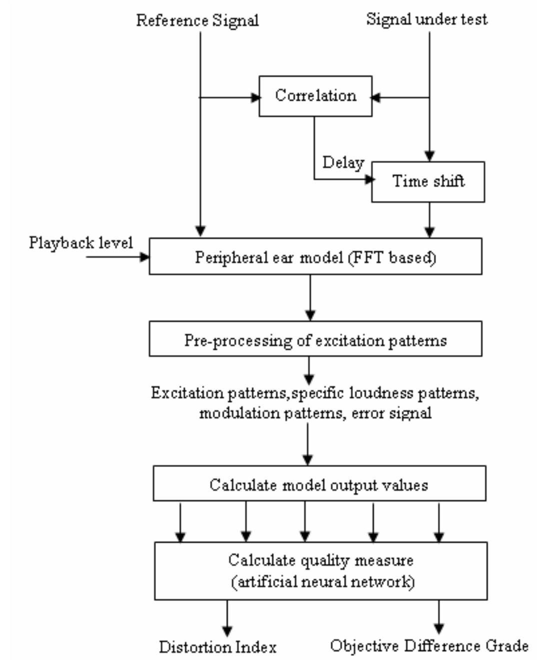



Fig. 4 Generic block diagram of MATLAB PEAQ version „basic" implementation

The first step of the program is devoted to a very important operation in a real measurement system: the time-aligning of input signals. This operation is essential, viewing the fact that coded version of the reference file presents usually an important delay in comparison with the original. The calculus of correlation function between the two acoustic signals reveals possible disalignements that are adjusted by a simple time-shifting of the test signal.

The input of the FFT-based ear model, the aligned reference and test signals, are cut into frames of about 0.042 s with an overlap of 50%. Each frame is transformed to the frequency domain using a Hann window and a short term FFT, and scaled to the playback level. It follows several intermediate steps (here referred as "*pre-processing of excitation patterns*"), the calculation of (mostly) psycho-acoustically based "*Model Output Variables"* (MOVs) and a mapping from the set of Model Output Variables to a single value representing the *basic audio quality* of the Signal under test.

The ITU-R Recommendation (ITU, 2001) imposes very strict conditions to verify if a program represents a proper implementation of the algorithm. In order to be a proper implementation, the program must reproduce for DI (Distortion Index) values of 16 pairs of WAV-files, the values given in Recommandation with a tolerance less than ±0,02 for all test items. Only 6 of 16 tests fulfilled the condition in the case of MATLAB implementation in Fig. 4, even if all the results showed a good conformance with the original values.

The same poor results at ITU-R BS.1387-1 conformance tests are also reported by Baumgarte and Lerch (2001), Kabal (2002) and Gottardi (2003). For this lack of conformance, the blame must be put on the unsatisfactory definition of the standard and the weak description of some of MOVs. There are also errors in the pseudo-code listed in the Recommendation. In defiance of this failure, our opinion is that the present MATLAB implementation of PEAQ algorithm stands as a very promising and effective tool in the perceptual assessment of audio materials.

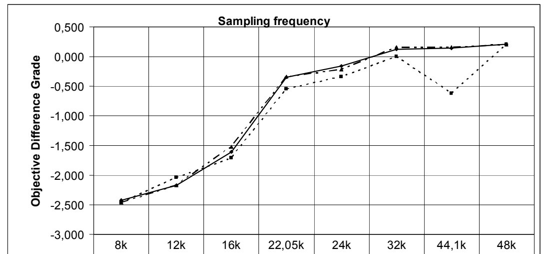
## 4. TEST RESULTS

In order to reveal possible areas of applications for the MATLAB implementation of PEAQ algorithm, some test procedures for the audio quality assessment were conceived. We intended to estimate the consequences of some usual digital processing techniques, including the encoding on various audio codecs, on the perceptual quality of an audio signal. An interesting issue can be also obtained by drawing parallels between our results and results obtained in similar subjective or objective tests (Stoll and Kozamernik, 2000; Salovarda *et all.*, 2004).

The choice of test material is crucial to the success of the tests and is far from being a simple matter. Three audio test stereo sequences, called *ref1*, *ref2* and *ref3*, were used. All of them are WAV-files (Microsoft Riff format), sampled at 48kHz, 16-bit PCM and have lengths between 14s and 19s. The first sequence, *ref1*, is pure speech, the others contain music. To carry out a PEAQ test, besides the reference file which is a WAV-file, the file containing the result of digital or codec processing must be finally converted to initial conditions: WAV format, 48kHz, 16-bit PCM.

### *4.1. Sampling frequency effect on the perceptual audio quality*

The reduction of the sampling rate of an audio material has very predictible effect on the quality of the audio file, viewing the direct dependence between the bandwidth of the audio signal and its sampling frequency. On the other side, it is obvious that for a high frequency limit of the human hearing of about 16kHz, the degradation of the perceptual quality of audio material is audible only at sampling rates under 32kHz. For the brief overall comparison, the final Objective Difference Grade (ODG) values for different sampling rates are shown in Fig. 5, separately for each of three audio files used. As it can be seen, the quality of the audio signal degrades significantly only at sampling rates under 22,05kHz. The content of audio file (speech or music) does not affect this process, as all files present very similar dependencies to sampling frequency.



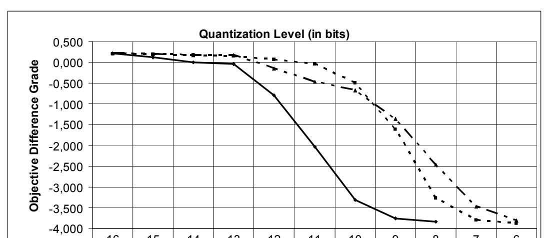

| | 8k | 12k | 16k | 22,05k | 24k | 32k | 44,1k | 48k |
|---|---|---|---|---|---|---|---|---|
| ref1 | -2,427 | -2,176 | -1,604 | -0,350 | -0,161 | 0,124 | 0,138 | 0,212 |
| ref2 | -2,461 | -2,036 | -1,714 | -0,547 | -0,341 | 0,002 | -0,622 | 0,209 |
| ref3 | -2,466 | -2,170 | -1,523 | -0,334 | -0,213 | 0,164 | 0,163 | 0,212 |

Fig. 5 The dependence of ODG values on sampling frequency.

### *4.2. ODG score dependence on the level of quantization.*

Quantization plays a major part in lossy data compression. In many cases, quantization can be viewed as the fundamental element that distinguishes lossy data compression from lossless data compression, and the use of quantization is nearly always motivated by the need to reduce the amount of data needed to represent a signal. In some compression schemes, like MP3 or Vorbis, compression is also achieved by selectively discarding some data, an action that can be analyzed as a quantization process (e.g., a vector quantization process) or can be considered a different kind of lossy process.

Initially, the reference WAV-file is quantized with 16 bits which can be one of 65,536 ($2^{16}$) possible values per sample. We reduced successively, bit by bit, the level of quantization from 16 bits to 6 bits. The results of PEAQ algorithm runnings on each of three reference files are shown in Fig. 6.




| | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ref1 | 0,212 | 0,121 | 0,000 | -0,044 | -0,799 | -2,033 | -3,308 | -3,758 | -3,832 | | |
| ref2 | 0,209 | 0,194 | 0,179 | 0,146 | 0,073 | -0,035 | -0,493 | -1,618 | -3,268 | -3,801 | -3,877 |
| ref3 | 0,212 | 0,198 | 0,181 | 0,179 | -0,151 | -0,474 | -0,677 | -1,362 | -2,470 | -3,469 | -3,807 |

Fig. 6 ODG values dependency on the level of quantization used to represent the audio signal.

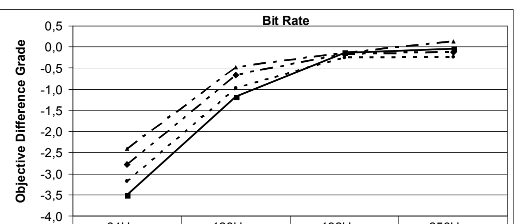
Some observations regarding Fig. 6 are obvious. For levels of quantization higher than 12-bits, the audio material quality is, as a matter of fact, unaffected by truncation. Next, the ODG score of file *ref1*, depreciate rapidly in contrast with the evolution of *ref2* and *ref3* files which continue to maintain a very high ODG score. For instance, at the level of 10-bits, the ODG score of *ref1* is -3,308, meaning a very annoying degradation of the audio material, while the ODG score of *ref2* and *ref3* reflect only a slightly perceptible degradation: -0,493 and -0,697. A possible explanation for this behaviour stands in the content of *ref1* file: pure speech, while *ref2* and *ref3* files contain music.

The quantization operation is a good example of nonlinear memoryless operation on each signal sample. At very low levels of quantization, *i.e.* lower than 8 bits, the quantization noise becomes more and more audible.

### *4.3. Comparison of usual audio codecs*

The PEAQ algorithm represents a very important resource in the objective perceptual quality comparison of audio codecs. Reliable results for this target means on the one hand to achieve the measurements in identical or equivalent conditions and on the other hand to perform the tests on a as large as possible database of audio files. These conditions were not have been fully fulfilled, viewing especially the limited number of test files (*ref1*, *ref2* and *ref3*). However the results submitted next, represent an average made on these three items measurements.

The codecs used are standard MP3 (or MPEG 1 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), OGG Vorbis (free codec from Xiph.org, different from MPEG 1 and 2 standards) and WMA 9.1 (Windows Media Audio 9.1 developed by Microsoft). 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.




| | 64kbps | 128kbps | 192kbps | 256kbps |
|---|---|---|---|---|
| mp3 | -3,506 | -1,179 | -0,146 | -0,038 |
| aac | -3,183 | -0,975 | -0,255 | -0,243 |
| ogg | -2,411 | -0,485 | -0,145 | 0,125 |
| wma | -2,774 | -0,661 | -0,163 | -0,120 |

Fig. 7 ODG values for 4 different codecs.

The codecs were tested at 4 usual bit rates, in all cases the sampling frequency being the same as for the reference files: 48kHz. This means that a possible resampling operation effect is excluded from measurement. For the brief overall comparison the averaged ODG values for 4 different bit rates are shown in Fig. 7.

Some conclusions can be drawn from Fig. 7. It is visible that all codecs act similarly at higher (≥192kbps) bit rates – the differences are minimal with the unexpected exception of AAC. On the lower bit rates (≤128 kbps), used extensively in *Internet Audio* and *Digital Broadcasting Radio*, 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 WMA is on the second position**.** The poor results of MP3 in these cases were a surprise.

From the results in Fig. 7 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, 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 codec in particular.

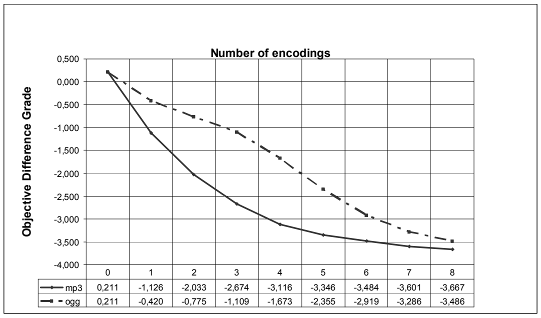
### *4.4. PEAQ evaluation of repeteated encodings effect*

In professional music studios, audio compression can happen at various stages between initial performance and final playback, or deployment in an audio testbed (Reiss and Sandler, 2004). Processing and transmission operations, change of formats and/or bit rates, all combine to create multiple cycles of decoding, processing, and re-encoding of the audio content. The quantization noise from each cycle accumulates and leads to a progressive drop in audio quality. The resultant distortion quickly becomes audible and becomes more and more problematic with each subsequent generation.

The effect of multiple encodings on perceptual audio quality degradation was measured for two most popular codecs: MP3 and OGG Vorbis. As vehicles of tests were used the same three test files used in all the tests. A single bit rate was used in all the encodings and for both codecs: 128kbps. All the encodings used a single sampling frequency: 48kHz. These identical conditions give the opportunity to compare the performance of codecs in the case of repeated encodings.

The result of PEAQ evaluation of repeated encodings effect for MP3 and OGG Vorbis codecs is presented in Fig. 8. As far as 8 successive encodings were made for each of three *ref* files in the above specified conditions. Each point of two curves is the average of these individual tests. OGG Vorbis free codec from Xiph.org has a much better behaviour than its rival:

the very known and popular format MP3. For instance, after two encodings if the perceptual quality of OGG files degraded just slightly (ODG=-0,775), the ODG score of MP3 files reveals an annoying degradation (ODG=-2,033).




| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| mp3 | 0,211 | -1,126 | -2,033 | -2,674 | -3,116 | -3,346 | -3,484 | -3,601 | -3,667 |
| ogg | 0,211 | -0,420 | -0,775 | -1,109 | -1,673 | -2,355 | -2,919 | -3,286 | -3,486 |

Fig. 8 Effect of multiple encodings on audio quality for MP3 and OGG audio formats.

To prevent audio quality deterioration of audio files by multiple encodings, very simple restrictions can be applied. First of all, the number of these operations must be kept to a minimum. Then, the repeated encodings must not affect the sampling frequency of the audio clip. Finally, it is better to perform all studio processings using the popular, well-understood and uncompressed format WAV.

## 5. CONCLUSIONS

This study is concerned with objective evaluation of perceived audio quality of audio signals. Sound quality is assessed using PEAQ (Perceptual Evaluation of Audio Quality) algorithm, and this greatly reduces the time-consuming efforts involved in listening tests.

The first target of the paper is to give a MATLAB implementation of ITU-R Recommandation BS.1387-1 which describes PEAQ algorithm. Unfortunately, the description of BS.1387-1 is inadequate by itself to allow for a conforming implementation. It is the reason why, as in other reported implementations, our MATLAB program failed to fulfill the conformance tests with the standard. Nevertheless, the present MATLAB implementation of PEAQ algorithm stands as a very promising and effective tool in the perceptual assessment of audio materials.

The last part of the paper is devoted to the development of some test procedures for the audio quality assessment of audio signals. The results permit the estimation of the consequences of some usual digital processing techniques, including the effect of simple or repeated encodings on various audio codecs, on the perceptual quality of an audio signal.

## REFERENCES

Baumgarte, F. and Lerch, A. (2001) *Implementation of Recommendation ITU-R BS.1387,* Delayed Contribution, Document 6Q/18-E, Feb. 2001.

Gottardi G. (2003) *Interpretazione e implementazione in “ANSI C” della raccomandazione ITU-R BS.1387-1,* Ancona, Universita Politecnica delle Marche

Kabal, P. (2002). *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. TSP Lab Technical Report, Dept. ECE, McGill Univ. May, 2002.

ITU. (2001). *ITU-R BS.1387-1*, *Method for the Objective Measurements of Perceived Audio Quality*, Geneva: International Telecommunication Union.

Reiss, J. and Sandler, M. (2004) Audio Issues in MIR evaluation. *Proc. 5’th International Conference on Music Information Retrieval*, pp. 28-33, Barcelona, Spain, October 10-14, 2004.

Salovarda, M., Bolkovac, I. and Domitrovic, H. (2004) Comparison of Audio Codecs using PEAQ algorithm, *6’th CUC Conference Proceedings*, Zagreb, Croatia, 27.9-29.9 2004.

Stoll, G. and Kozamernik, F. (2000). EBU listening tests on Internet audio codecs. *EBU Technical Review*, no. 283, pp. 1- 24, June 2000.

Thiede T., Treurniet W. C., Bitto R., Schmidmer C., Sporer T., Beerends J. G., Colomes C., Keyhl M., Stoll G., Brandenburg K., and Feiten B. (2000). PEAQ – The ITU Standard for Objective Measurement of Perceived Audio Quality. *J. Audio Eng. Soc.,* vol. **48**, pp. 3–29, 2000.