Tools for Video Analysis
1566 La Pradera Dr
Campbell, CA 95008
www.videoclarity.com
408-379-6952

Application Note – PEAQ Audio Objective Testing in ClearView

Video Clarity, Inc.

General Description

The PEAQ measurement is now part Video Clarity's ClearView system. ClearView systems provide several proprietary tools for audio measurement and now PEAQ provides an international standard method for measuring objective audio quality.

PEAQ is an implementation of Recommendation BS-1387, an international standard developed by the ITU-R (International Telecommunication Union – Radiocommunication Sector) [1]. Video Clarity brings this method to ClearView as certified by the Communications Research Centre, Canada, one of the major contributors to the development of this standard.

BS-1387 describes an objective method for measuring perceived audio quality based on a computational model of the human auditory system [2]. The acronym for the measurement model is PEAQ (Perceptual Evaluation of Audio Quality).

The PEAQ model can be viewed, as shown in Figure 1, as an "electronic ear" which compares an impaired signal under test with a reference signal and generates a mean quality rating.

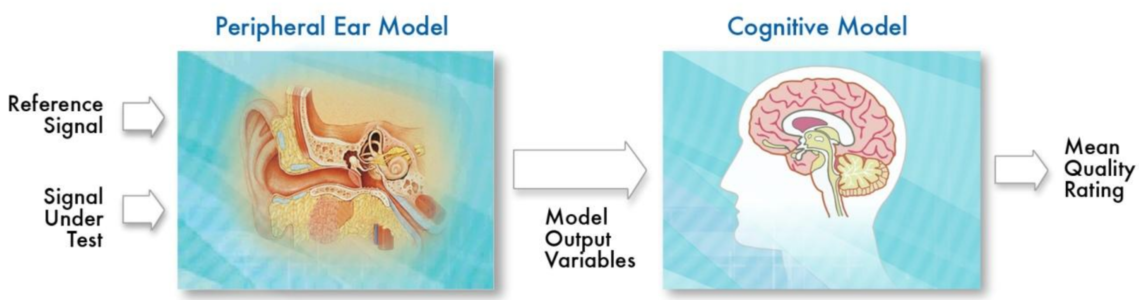

Figure 1 – A simplified block diagram of the PEAQ model

The PEAQ model was calibrated to output scores correlated with the mean opinion scores of a panel of expert listeners taking part in a formal subjective test [2], [3]. In this sense, PEAQ can be viewed as an "average expert listener".

PEAQ is optimized to measure the perceptual quality of wideband (20kHz) audio codecs and has been successfully used to measure the quality of other wideband audio devices such as analog-to-digital converters, digital-to-analog converters and sample rate converters [4].

The mean quality rating is either scored as a 0 to -4 or 1 to 5. The score is an Objective Differential Grade (ODG) which represents the overall severity of the impairments in the test signal and its meaning is defined on the rating scales of Figure 2 (page 3).

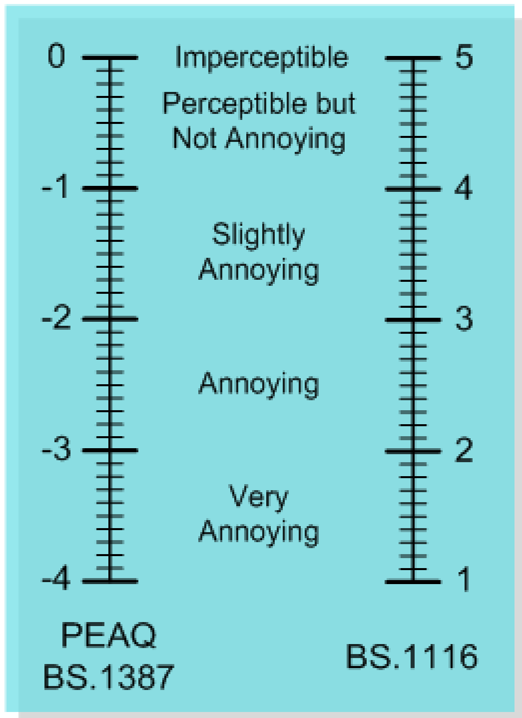

Figure 2 – PEAQ impairment scales

Applications of PEAQ

PEAQ can be used in a number of different applications shown in Table 1. More information can be found on these applications in [1] and [2].

	Application	**Description**
1	Assessment of implementation	A procedure for characterizing different implementations of audio processing equipment, in many cases audio codecs
2	Perceptual quality line-up	A continuous process to monitor an audio transmission in service
3	Off-line monitoring	A detailed analysis of a piece of equipment or a circuit
6	Codec development	A procedure to optimize the cost and performance of a transmission network under given constraints
7	Network planning	A procedure to optimize the cost and performance of a transmission network under given constraints
8	Aid to subjective assessment	A tool for identifying critical material to include in a listening test

Table 1 - List of applications for the PEAQ model

Measuring Audio Quality with PEAQ using ClearView

Measuring Audio Quality with PEAQ using ClearView

There are several operational examples as to how ClearView can be used to measure audio quality with PEAQ. See Figure 3 for a workflow example.

1. Dual input. Dual input is the mode where the two AV signals to be compared (i.e. original reference signal and impaired version) are both input as SDI, HDMI or Analog AV and recorded in ClearView. The two sequences are then automatically aligned and a PEAQ quality measurement is performed on the audio channels associated with the video program.

2. Audio reference file load, then play/record. The original reference sequence and processed sequence to be tested may be input from file via directly connected external hard drive or over IP network. A copy of the reference audio file is first attached to a video sequence and loaded into a ClearView library. This resulting ClearView AV sequence can then be played through a processing chain or device under test while a recording is made with the system's input interface (see Figure 3). As in example 1 above, the reference and recorded sequences are then automatically aligned and a PEAQ measurement is performed on the audio channels.

3. Audio file load and test. In this example, as in #2 above, an original reference audio file is loaded into a ClearView library and attached to a video sequence. The same reference audio file is then put through the target processing. The resulting audio file is then loaded into ClearView and coupled with a copy of the same video sequence as the reference so that an alignment and PEAQ measurement can be made.

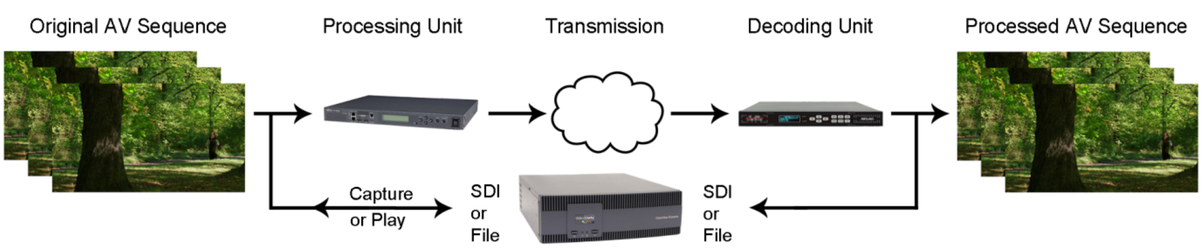

Figure 3 – ClearView workflow

Basic versus Advanced PEAQ models

The PEAQ model described in ITU-R Recommendation BS-1387 includes two versions: an FFT-based ear model labeled as the Basic Version and a filterbank-based ear model known as the Advanced Version [1] [2]. Figures 4 and 5 from [2] show no significant difference in performance (i.e. correlation between objective (ODG) and subjective (SDG) ratings) between the Basic Version and the Advanced Version over a database of 812 audio sequences subjectively assessed in 9 different formal subjective test experiments. The Basic Version is implemented in PEAQ for ClearView because it provides the same performance and is less complex than the Advanced Version.

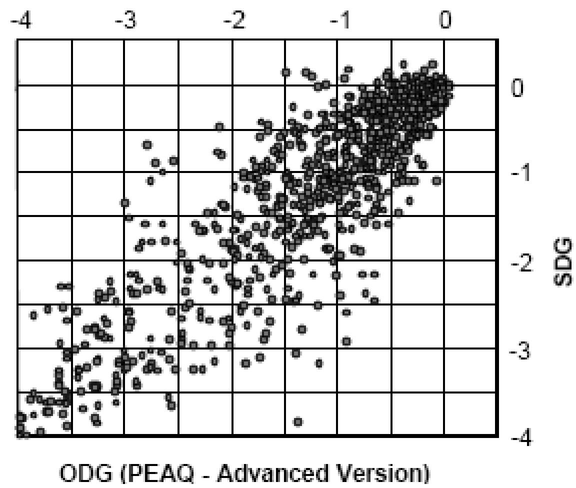

Figure 4 - PEAQ Advanced model prediction vs.
Listening test results for all available test data [2]

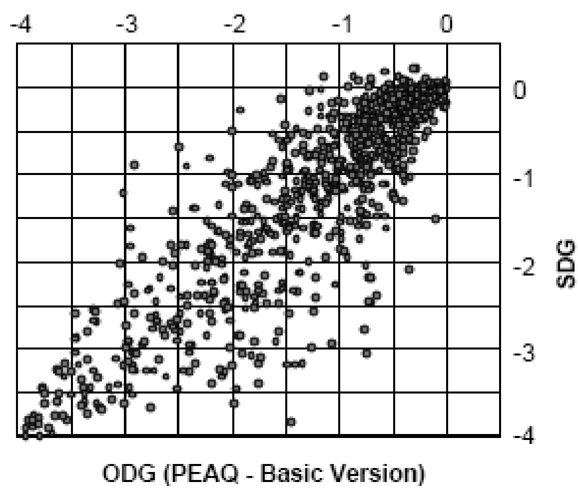

Figure 5 - PEAQ Basic model prediction vs.
Listening test results for all available test data [2]

Increasing the Reliability of PEAQ ODG's

Subjective Test
In a formal subjective test of audio codecs conducted at the Communications Research Centre in 1997 [5], 21 expert listeners evaluated the quality of 8 audio items processed by 17 systems. The systems consisted in six different audio codecs (labeled U, V, W, X Y and Z) operated at bit rates ranging from 64 to 192 kbps per stereo pair as shown in Table 3.

Label	Bitrates (Kbps/stereo channel)
Codec U	64, 96, 128, 160
Codec V	128
Codec W	128, 160, 192
Codec X	128, 160, 192
Codec Y	96, 128

Table 3 Codec labels and bit rates [3]

By averaging over the 21 listeners, the subjective data set is reduced to 136 mean opinion scores (8 audio items x 17 systems), labeled here as Subjective DiffGrades or SDGs. Each of the 136 coded (i.e. distorted) audio sequences used in the listening test was compared to its uncoded (i.e. reference) counterpart by the PEAQ model which generated an ODG for each of those coded sequence.

The performance of the PEAQ model was evaluated in two ways [2]:

- First, the objective and mean subjective quality measurements are compared for each critical audio item.
- Second, the objective and subjective overall system quality measurements were compared by averaging codec quality measurements over the eight critical audio items.

Comparison by Audio Item

Figure 6 shows the relationship between the mean SDG and the ODG for the 136 audio items. The linear correlation between these variables is 0.85, and the slope of the regression line is 0.79. Perfect correspondence between the objective measurements and the subjective quality ratings was not achieved since all of the data points do not fall on the positive diagonal. However, the objective measurements agree reasonably well with the subjective quality grades.

Comparison by System

In the original subjective study [5], the overall subjective quality of a particular system was defined as the average of the mean SDGs for the eight audio items processed by that system. The corresponding overall objective quality measurement was obtained by averaging the ODGs for the same eight audio items.

Figure 7 shows the relationship between the 17 overall mean SDGs and ODGs. The slope of the regression line is 0.95 and the linear correlation between the variables is 0.97. The high correlation combined with a slope approaching unity indicates that the overall objective quality of a number of critical audio items is an accurate and unbiased indicator of the perceived quality of a device. The correlation between overall mean ODG and overall mean SDG shown in Figure 7 is much stronger than the correlation between ODG and mean SDG for individual items shown in Figure 6. This can be understood as a consequence of averaging over subsets of audio items.

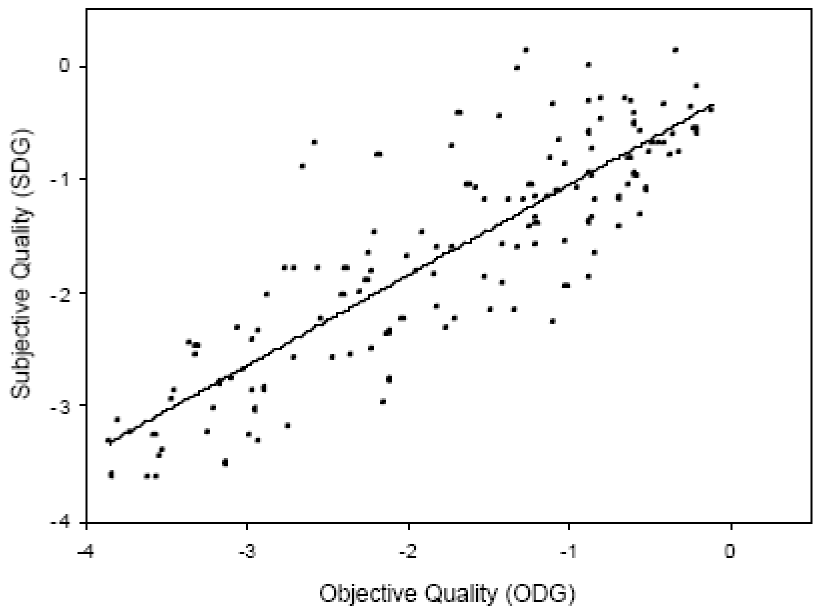

Figure 6 Correlation of mean item ODG with SDG (r = 0.85) [3]

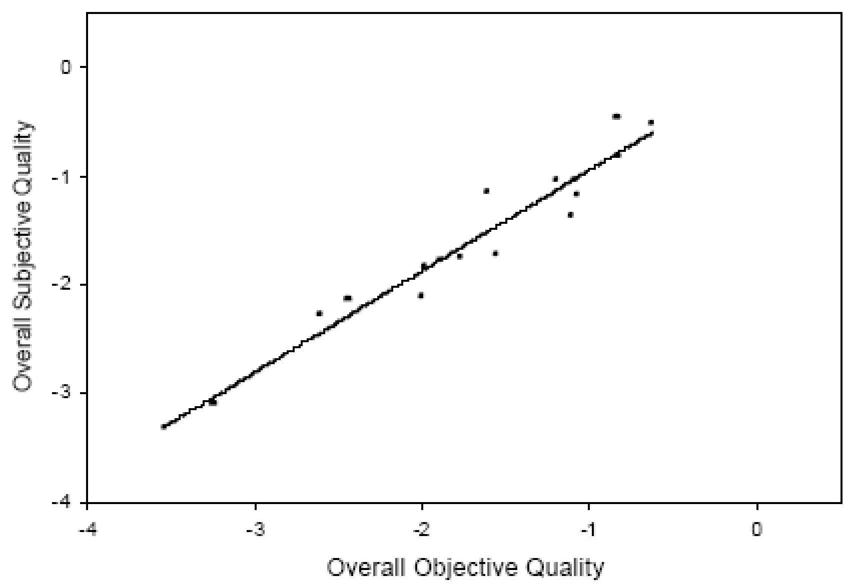

Figure 7 Correlation of overall mean system ODG with SDG (r = 0.97) [3]

The above results show that the PEAQ model is a reliable tool for obtaining quality measurements of perceptual audio codecs. The model generated objective quality measurements for coded audio signals previously used in a listening test [1]. The objective quality measurements averaged over audio items were found to be statistically indistinguishable from similarly averaged subjective quality ratings. Further, the correlation between the means of the SDG's and ODG's over several audio items was considerably stronger than that for individual items.

Therefore, when an overall measurement of system quality is desired, simply averaging the objective quality measurements over a number of different audio test sequences (6 to 10) was shown to enhance the reliability of the objective measurements.

Audio Test Signals

When testing the audio quality of devices, it is common to use critical audio test signals, i.e. test signals which are difficult to process by the device under test. This is particularly true for audio codecs. The following table provides a list with a subset of test signals that were used during the verification procedure that led to ITU-R Recommendation 1387-1 [1]. The type of artifacts, which these signals typically reveal due to low bit-rate coding, is also indicated.

No.	Item	Remarks
1	Castanets	1
2	Clarinet	2
3	Claves	1
4	Flute	2
5	Glockenspiel	1 & 2 & 5
6	Harpsichord	1 & 2 & 4
7	Kettle drum	1
8	Marimba	1
9	Piano Schubert	2
10	Pitch Pipe	4
11	Ry Cooder	2 & 4
12	Saxophone	2
13	Bag Pipe	2 & 4 & 5
14	Speech Female Engl.	3
15	Speech Male Engl.	3
16	Speech Male German.	3
17	Snare Drums	1
18	Soprano Mozart	4
19	Tambourine	1
20	Trumpet	2
21	Triangle	1 & 2 & 5
22	Tuba	2
23	Susanne Vega	3 & 4
24	Xylophone	1 & 2

Table 4 List of possible test signals for use with PEAQ

Remarks:
1) Transients: pre-echo sensitive, smearing of noise in temporal domain.
2) Tonal structure: noise sensitive, roughness.
3) Natural speech (critical combination of tonal parts and attacks): distortion sensitive, smearing of attacks.
4) Complex sound: stresses the Device Under Test.
5) High bandwidth: stresses the Device Under Test, loss of high frequencies, program-modulated high frequency noise.

The test signals given in Table 4 are from the EBU SQAM compact disc [6]. They can be used free of copyright only for measuring purposes together with PEAQ. Clearance of copyright has to be obtained for all sequences, mainly from the EBU (European Broadcast Union).

The duration of a test signal for PEAQ should be about the same as if it were to be used in a listening test. This duration is typically in the order of 10 to 20 seconds. It is very likely that the critical part of the test signal, which reveals most of the artifacts, is limited to only a short part of the duration. Considering these buffer lengths and the time constants present in the PEAQ measurement method, the duration of each single test item in a sequence must be more than 500 milliseconds.

The Signal Under Test and the Reference Signal must be synchronized in time for PEAQ to make a valid quality measurement. The ClearView analyzer system has a built-in automatic synchronization feature to ensure valid measurements.

References

[1] ITU-R Rec. BS-1387-1, "Method for Objective Measurements of Perceived Audio Quality", International Telecommunications Union, Geneva, Switzerland (2001)

[2] T. Thiede, W.C. Treurniet et al., "PEAQ - the ITU Standard for Objective Measurement of Perceived Audio Quality", Journal of the Audio Engineering Society, vol. 48, Number 1/2, Jan/Feb 2000.

[3] W.C Treurniet and G. Soulodre, "Evaluation of the ITU-R Objective Audio Quality Measurement Method", Journal of the Audio Engineering Society, vol. 48, No. 3, March 2000

[4] E. Benjamin, "Evaluating Digital Audio Artifacts with PEAQ", Convention Paper 5711, 113th Convention of the Audio Engineering Society, Los Angeles, USA, 5-8 Oct. 2002

[5] G.A. Soulodre, T. Grusec, M. Lavoie and L. Thibault, "Subjective Evaluation of State-of-the-Art Two-Channel Audio Codecs", Journal of the Audio Engineering Society, vol. 46, No. 3, pp. 164-177, March 1998

[6] EBU document Tech. 3253 (1988): Sound quality assessment material. Recordings for subjective tests – Users' handbook for the EBU-SQAM Compact Disc, European Broadcast Union

The major statistical sections of this application note were republished by permission of the Communications Research Centre Canada.