# EVALUATING LOW BITRATE SCALABLE AUDIO QUALITY USING ADVANCED VERSION OF PEAQ AND ENERGY EQUALIZATION APPROACH

*Rahul Vanam and Charles D. Creusere*

Klipsch School of Electrical & Computer Engineering, New Mexico State University, Las Cruces, NM 88003-8001, USA
{rahulv,ccreuser}@nmsu.edu

## ABSTRACT

The ITU-R BS.1387-1 gives a method for objective measurement of perceived audio quality known as PEAQ (Perceptual Evaluation of Audio Quality). This algorithm has been developed for measuring the quality of mid and high quality audio. In this paper we show that the Advanced version of the PEAQ performs poorly when compared to the previously developed Energy Equalization Approach (EEA) for evaluating quality of low bitrate scalable audio. We also show that including Energy Equalization parameter as one of the Model Output Variables (MOVs) of the Advanced version will improve its performance significantly; the performance of this modified version is superior to that of EEA.

## 1. INTRODUCTION

Lossy audio compression algorithms are popular as they provide higher compression compared to their lossless counterparts. Most of these algorithms take advantage of the perceptual characteristics of the human auditory system like absolute hearing threshold, simultaneous masking, spread of masking along the Basilar membrane and temporal masking [1]. To evaluate the quality of such compressed audio, subjective listening tests are required. Since these are often time consuming and impractical, an objective measurement method based on the perceptual model of the human ear is preferred. Many algorithms have been proposed for objective measurement of audio quality [2]-[6] and their best features have been combined into a single measurement method brought out as a recommendation by the International Telecommunication Union (ITU) i.e. ITU-R BS.1387-1 [7]. This recommendation includes two versions–the Basic and Advanced versions–which tradeoff accuracy and speed. The Basic version includes a Fast Fourier Transform (FFT)-based ear model whereas the Advanced version includes an ear model based on both the FFT and a Filter bank. To be more specific, the Advanced version generates 5 psychoacoustically based Model Output Variables (MOVs) which include parameters for distortion loudness, changes in modulation, linear distortion and harmonic structure of the error [7]. The MOVs are mapped to a single quality measure called the Objective Difference Grade (ODG) using an artificial neural network.

Most audio compression standards ensure perceptual transparency at high to mid bitrates. The method for assessment of its quality using subjective testing is done according to ITU-R BS.1116 [8] and the objective measurement is done using the PEAQ. More recent standards like Motion Pictures Experts Group-4 (MPEG-4) support scalable audio compression that encodes audio data at a higher bitrate and decodes it at bitrates less than or equal to the original bitrate. Objective quality measurement of low bitrate scalable audio using PEAQ has been found to be poor for the Basic version [9]. In this paper we show that the Advanced version also performs poorly for high impairment audio using subjective test data and, further, that EEA is superior to it. We also show that incorporating the EEA into the Advanced version will improve its performance considerably.

## 2. SUBJECTIVE TESTING

In our subjective and objective audio quality measurements, we use codecs (encoder/decoder) from MPEG-4 family namely Bit Slice Arithmetic Coding (BSAC), Transform Weighted Interleaved Vector Quantization (TwinVQ) and Advanced audio coder (AAC). BSAC is a scalable codec which is a variant of AAC, TwinVQ is a nonscalable codec that is known to perform well at low bitrates, and AAC is nonscalable, performing best at higher bitrates. We use the audio sequences and follow the subjective test methodology as discussed in [9]. Specifically, we work with seven monaural sequences, two of which are from MPEG-4 test set and the rest from various classical and popular music sources. Each audio sequence is encoded and decoded using the above mentioned codecs. For BSAC, we encode the audio at 64 kbits/s (kb/s) and decode it at 32 kb/s and 16 kb/s. For

Research sponsored by the NSF, grant # CCR-0133115.

TwinVQ, we encode and decode at 32 kb/s and 16 kb/s, and for AAC we encode and decode at 32 kb/s. In addition to 16 kb/s BSAC, we use another variant of BSAC for which the original audio is low-pass filtered to 6 KHz prior to encoding at 64 kb/s and decoding at 16 kb/s respectively.

Since we perform subjective tests for high impairment audio, Comparison Category Rating (CCR) approach is followed [10]. The CCR rates two audio sequences on a scale of -3 to 3. A score of 0 indicates that the two algorithms are equivalent and a score of 3 indicates that the first is ‘much, much better’ compared to the second. In [9] it is shown that TwinVQ at 16 kb/s has significantly better perceptual quality compared to BSAC at 16 kb/s. Pre-filtering the BSAC-compressed audio improves the quality of reconstructed audio by a small amount. Furthermore, BSAC at 32 kb/s is found to be equivalent to that of nonscalable algorithms [9].

## 3. PERFORMANCE COMPARISON

In this section, we compare the performance of the Advanced ITU metric to that of the EEA. We also evaluate the Advanced version with and without Energy Equalization parameter as one of its MOVs. A comparison is made based on the correlation coefficient that is obtained from subjective and objective test data and the robustness of the metric in predicting the audio quality.

### 3.1. Advanced ITU metric versus Energy Equalization

The EEA has been discussed in detail in [9] where it was shown to achieve superior performance when compared to the Basic ITU metric for measuring the quality of highly impaired audio. In this section we compare performance of Advanced version to that of Energy Equalization algorithm.

In the EEA, we first compute the energy of an encoded/decoded audio signal in the frequency range 2.2 to 4.3 KHz. We then compute truncation threshold $T_{kn}$ for each codec $k$ and audio sequence $n$ such that the energy of original uncoded audio truncated by a threshold value $T_{kn}$ equals the energy of the test audio. The comparison between two codecs is performed in differential manner since subjective comparison is itself differential. The optimal predictor based on truncation threshold is determined by solving the linear equation

$$\mathbf{a}\mathrm{x} = \mathbf{p} \qquad (1)$$

for a scalar $x$ where **a** is a column vector containing differential threshold data and **p** is a column vector containing the subjective test data. The vector **p** and **a** in our case are $14x1$ vectors since we consider two comparisons (namely BSAC at 16 kb/s vs. TwinVQ at 16kb/s and BSAC at 16 kb/s vs. BSAC at 16 kb/s with pre-filtering) for 7 audio sequences. The vector **a** is scaled to have values in the range (0,3] which corresponds to the absolute range of our subjective data. Also **a** is differential. The least square solution to (1) is given by

$$\hat{x} = (\mathbf{a}^{\mathrm{T}}\mathbf{a})^{-1}\mathbf{a}^{\mathrm{T}}\mathbf{p}. \qquad (2)$$

The solution from (2) is plotted in Fig.1 and it is observed that the predictor $\hat{x}$ has a slope close to 1.0. This indicates that the difference in thresholds can be used directly as a metric for measuring relative audio quality between two codecs. Testing for the robustness of the predictor is done by successively eliminating each of the 14 test cases from (1) and computing the predictor $\hat{x}$ using (2) for the modified system. The optimal predictor is then applied to the data point that was not used in the design process. Figure 2 shows that the squared error is less than 1.0 except for two cases. This indicates that the training set depends upon audio sequences whose corresponding squared error is greater than 1.0. The robustness may improve if a larger number of audio sequences of different types are included in our training set.

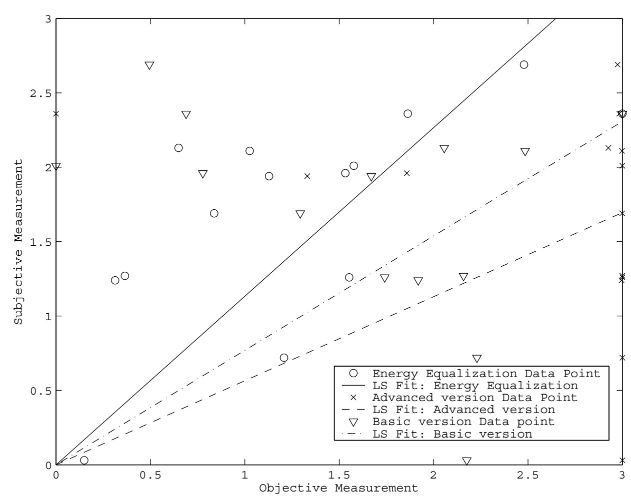


**Figure 1**. Least square fit of objective quality measure versus subjective data.

We now consider the Objective Difference Grade values $ODG_{kn}$ of the Advanced version, computed for each audio sequence $n$ and encoded and decoded with codec $k$. From these values, we obtain a vector containing difference in ODG values corresponding to pairs of reconstructed audio sequences i.e., $\mathbf{a} = [(ODG_{k_1 1} - ODG_{k_2 1}), (ODG_{k_1 1} - ODG_{k_3 1}),\ldots, (ODG_{k_i n} - ODG_{k_m n})]^{\mathrm{T}}$. The predictor $\hat{x}$ for the linear system (1) with redefined **a** is determined using (2). Using the redefined vector **a** we obtain correlation coefficient of 0.3259 versus 0.6694 for the EEA truncation threshold. Note that the correlation coefficient is a measure of the degree of linear relationship present between two variables [11]. For variables having positive relationship, the correlation coefficient is 1.0 for perfect linear relationship and 0.0 for no linear relationship. This indicates a closer relationship between objective and subjective measurements for EEA than for the Advanced version of PEAQ. The correlation coefficient for Basic version (EAQUAL - Evaluation of Audio Quality, version 0.1.3 alpha [12]) is 0.3655 indicating better performance when compared to the Advanced version. From Fig 1, we observe that the EEA performs better than both the Advanced and Basic versions of the ITU recommendation. It is interesting to observe that the Advanced version cannot differentiate between BSAC at 16 kb/s and BSAC at 16 kb/s with pre-filtering. This is indicated by the Advanced version data points that are close to the Y axis. However, subjective tests indicate that pre-filtered BSAC at 16 kb/s sounds 'a little better' compared to BSAC at 16 kb/s [9].

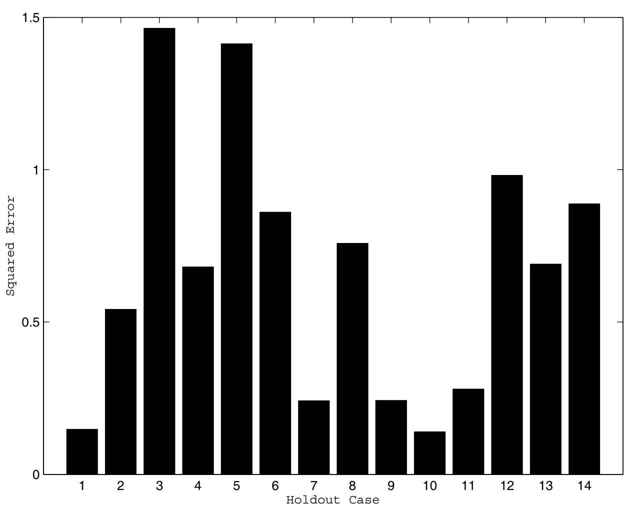


**Figure 2.** Squared error in Energy Equalization approach when numbered case is not used to design the predictor.

### 3.2. Advanced version with and without Energy Equalization MOV

In [9] it was concluded that relative performance of two codecs could be directly calculated from the difference of their truncation threshold. We therefore consider the possibility of including the truncation threshold as an additional MOV in the Advanced version of the ITU metric in order to improve its performance on highly impaired audio. These 6 MOVs are mapped to a single quality measure, and this is done using a simple linear equation given by

$$d = \mathbf{m}^T\mathbf{w} \tag{3}$$

where $d$ is the ODG value, **m** is a $6x1$ vector containing 6 MOVs and **w** is a vector $[w_1, w_2, ..., w_6]^T$ containing weights for the MOVs. We compute the weights by solving the following linear equation

$$\mathbf{Aw} = \mathbf{p} \tag{4}$$

where **A** represents an $mx6$ matrix $\begin{bmatrix} (MOV_{1,1} - MOV_{2,1}),..,(MOV_{1,6} - MOV_{2,6}) \\ : \quad : \quad : \\ (MOV_{m-1,1} - MOV_{m,1}),...,(MOV_{m-1,6} - MOV_{m,6}) \end{bmatrix}$

containing the difference of MOVs and **p** is a vector containing subjective test data. In our experiments, **A** is $14x6$ matrix, **w** is $6x1$ vector and **p** is $14x1$ vector. The system represents an over-determined system and the least square solution for the weights is given by

$$\hat{\mathbf{w}} = (\mathbf{A}^T \mathbf{A})^{-1}\mathbf{A}^T\mathbf{p}. \tag{5}$$

Once the weights are determined, we compute the ODGs for each audio sequence using (3). The ODG differences are computed and a linear equation is built as in (1). Its least squares solution $\hat{x}$ is determined using (2). The predictor is tested for robustness for perturbations in the training set using the method stated in the previous section. From Fig 4, we observe that the maximum squared error is less than 0.9, which is smaller than that for Energy equalization approach. The correlation coefficient for the modified Advanced version is found to be 0.8254, indicating that it has superior performance over EEA, ITU-basic and ITU-advanced metric. Figure 3 shows the least square fit for Advanced version with and without Energy equalization parameter as its MOV. The fit for the Basic version is also shown. Table 1 compares the correlation coefficient and slope parameters of the different objective measurement methods discussed in this paper. Clearly, the ITU-advanced metric with the Energy Equalization parameter as an MOV performs the best in measuring high impairment audio quality.

**Table 1.** Parameters from various objective measurement schemes.

| Methods for Objective measurement of audio quality | Correlation coefficient | Slope of its LS fit |
|---|---|---|
| Basic version (EAQUAL version alpha 0.1.3) | 0.3655 | 0.7694 |
| Advanced version | 0.3259 | 0.5651 |
| Energy Equalization approach | 0.6694 | ≈ 1.0 |
| Modified Advanced version | 0.8254 | ≈ 1.0 |

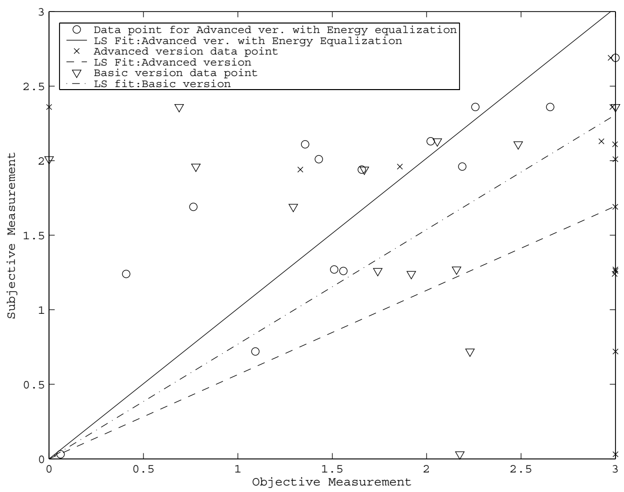


**Figure 3.** Least square fit of objective quality measure versus subjective data.

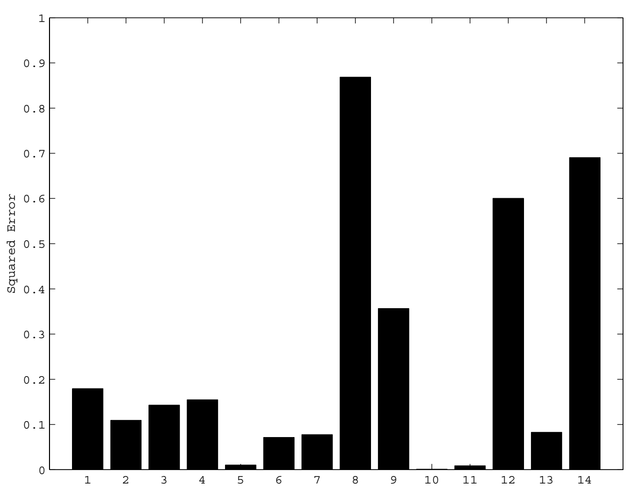


**Figure 4.** Squared error in the modified Advanced version when numbered case is not used to design the predictor.

## 4. CONCLUSIONS

In this paper we have shown that the Advanced version of PEAQ performs poorly for measuring low bitrate scalable audio quality compared to both the Basic version and the EEA. By including the Energy Equalization parameter as an additional MOV in the Advanced ITU metric, the performance is better than either the Basic ITU metric or the EEA alone. Since ITU-R BS.1534-1 [13] provides a method for subjective assessment of high impairment audio quality and is recent compared to the CCR approach, we plan to follow this recommendation for obtaining subjective test data in our future research. Also, the performance of Advanced version with and without Energy equalization parameter will be evaluated for the 32 kb/s audio data.

## 5. REFERENCES


[1] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings Of the IEEE*, Vol.88, No.4, April 2000.

[2] K. Brandenburg, "Evaluation of quality of audio encoding at low bit rates," *82nd AES Convention of the Audio Engineering Society,* preprint 2433, London, UK, February 1987.

[3] T. Thiede and E. Kabot, "New Perceptual Quality measure for the bitrate reduced audio," *100th Convention of the Audio Engineering Society,* preprint 4280, Copenhagen, Denmark, 1996.

[4] J. G. Beerends and J. A. Stemerdink, "A perceptual audio quality measure based on psychoacoustic sound representation," *J.Audio Eng. Soc.,* Vol. 40, pp.963-978, December 1992.

[5] B. Paillard, P. Mabilleau, S. Morissette and J. Soumagne, "PERCEVAL: Perceptual evaluation of the quality of audio signals," *J. AudioEng. Soc. ,*Vol.40, pp. 21-31, Jan./Feb 1992.

[6] M. P. Hollier, D. R. Guard, and M. O. J. Hawksford, "Objective perceptual analysis: Comparing the audible performance of data reduction schemes," *96th Convention of the Audio Engineering Society*, preprint 3797, Amsterdam, 1994.

[7] *Method for objective measurements of perceived audio quality,* Recommendation ITU-R BS.1387-1, Geneva, Switzerland, 1998-2001.

[8] *Methods for subjective assessment of small impairments in audio systems including multichannel sound systems,* Recommendation ITU-R BS.1116, Geneva, Switzerland, 1994.

[9] C. D. Creusere, "Quantifying perceptual distortion in scalably compressed MPEG audio," *37th Asilomar Conference on Signals, Systems and Computers,* Monterey, U.S.A, pp.265-269, November 2003.

[10] *Subjective performance assessment of telephone-band and wide-bandwidth digital codecs*, Recommendation ITU-R P.830, Geneva, Switzerland, 1996.

[11] A. L. Edwards, *An introduction to linear regression and correlation*, W. H. Freeman, New York, 1984.

[12] Source code for EAQUAL can be downloaded from the website: http://www.mp3-tech.org/programmer/misc.html.

[13] *Method for the subjective assessment of intermediate quality level of coding systems,* Recommendation ITU-R BS.1534-1, Geneva, Switzerland, 2001-2003.