

Scalable Perceptual Metric for Evaluating Audio Quality

Rahul Vanam

Dept. of Electrical Engineering
University of Washington

Charles D. Creusere

Klipsch School of Electrical and Computer Engineering
New Mexico State University



Background

- Because modern audio compression algorithms are optimized for the human auditory system, conventional objective like segmental signal-to-noise ratio are not effective
- This has forced researchers to rely upon human subjective testing in order to validate and compare different algorithms
 - Time consuming
 - Ill-suited to online implementation
 - The results are often difficult to repeat

Background

- Since the late 1980s, there has been a strong push to develop objective metrics capable of quantifying subjective audio quality
- This culminated in the development of ITU-R Recommendation BS.1387-1, called PEAQ
 - Contains a lower-complexity basic version and a more accurate advanced version



Background

- ***Problem:*** Both the basic and advanced versions of PEAQ are designed to evaluate the quality of mildly impaired audio
- Because we are interested in scalable audio compression, we would like an objective metric that is accurate over a wide range of audio impairments



Solution Framework

- In previous work, we found that an alternative metric, namely the Energy Equalization Approach (EEA), was far more accurate in characterizing the quality of highly impaired audio than either version of PEAQ
- In this paper, we combine EEA with PEAQ-advanced to create a metric that is fidelity scalable: i.e., that is accurate over a wide range of audio qualities.



Energy Equalization Approach

- **Idea:** Apply a truncation threshold to the original audio sequence, adjusting it until the energy of this sequence is the same as that of the reconstructed audio sequence
 - Mimics the process of band truncation that occurs in perceptual audio codecs



Energy Equalization Metric

Define:

- Energy of reconstructed audio

$$e_k = \sum_{i=0}^{total_blocks} \sum_{j=51}^{100} (\text{rec_spec}(i, j)_k)^2$$

GOAL: Select T so that: $e_T = e_k$

- Modified time-frequency spectrum

$$\mathbf{m_spec}(i, j)_{T_{kn}} = \begin{cases} \mathbf{o_spec}(i, j), & \text{if } |\mathbf{o_spec}(i, j)| \geq T_{kn} \\ 0, & \text{if } |\mathbf{o_spec}(i, j)| < T_{kn} \end{cases}$$

- Energy of modified spectrum

$$e_{T_{kn}} = \sum_{i=0}^{total_blocks} \sum_{j=51}^{100} (\mathbf{m_spec}(i, j)_{T_{kn}})^2$$

New Metric Design

- We combine the 'T' parameter generated by EEA with the five Model Output Variables (MOVs) that are already part of the PEAQ-advanced recommendation
 - Existing MOVs quantify the distortion loudness, the changes in modulation, the linear distortion, the harmonic structure of the error, and the noise-to-mask ratio
- A simple optimal linear weighting is used to fuse the MOVs into a single value



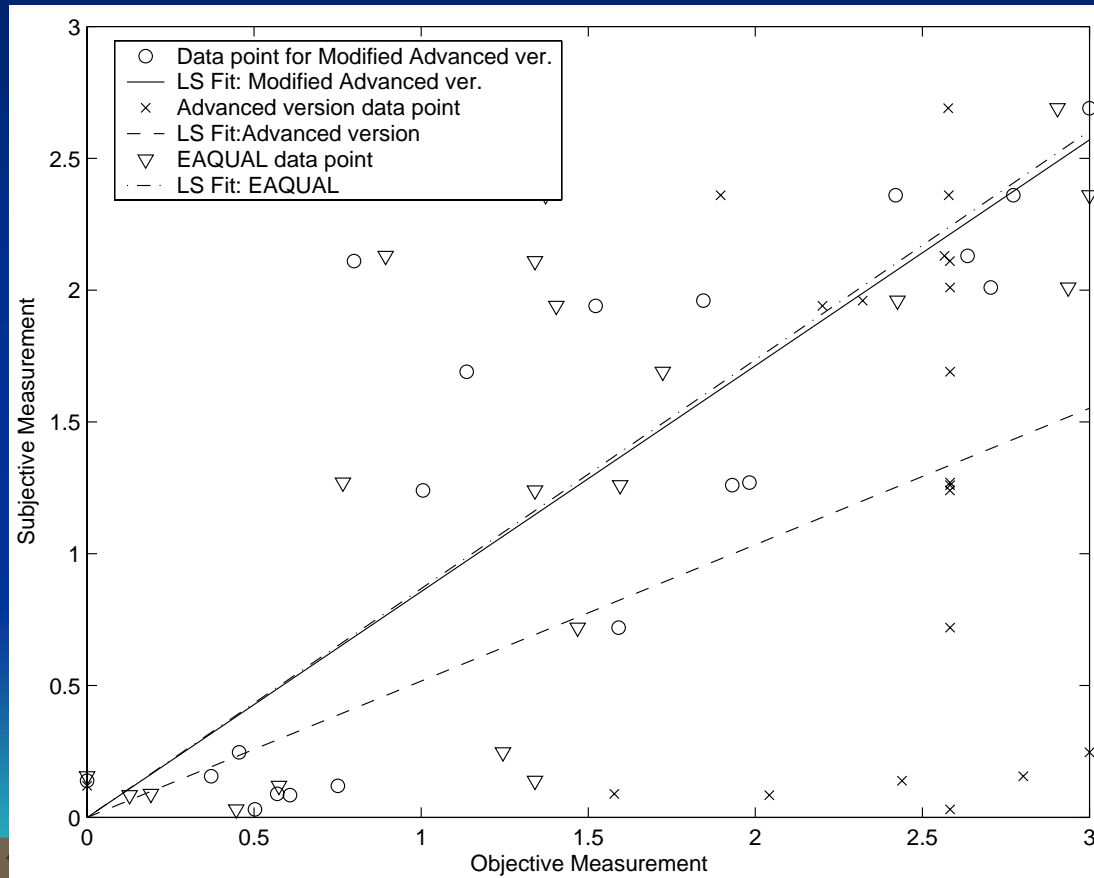
Subjective Test Data

- The data used to test and design the proposed objective metrics was collected using the Comparison Category Rating (CCR) approach
 - 20 test subjects
 - 7 different audio sequences
 - Encoded bitrates of 16 and 32 kb/s
 - Using MPEG4 codecs: AAC, BSAC, and TVQ



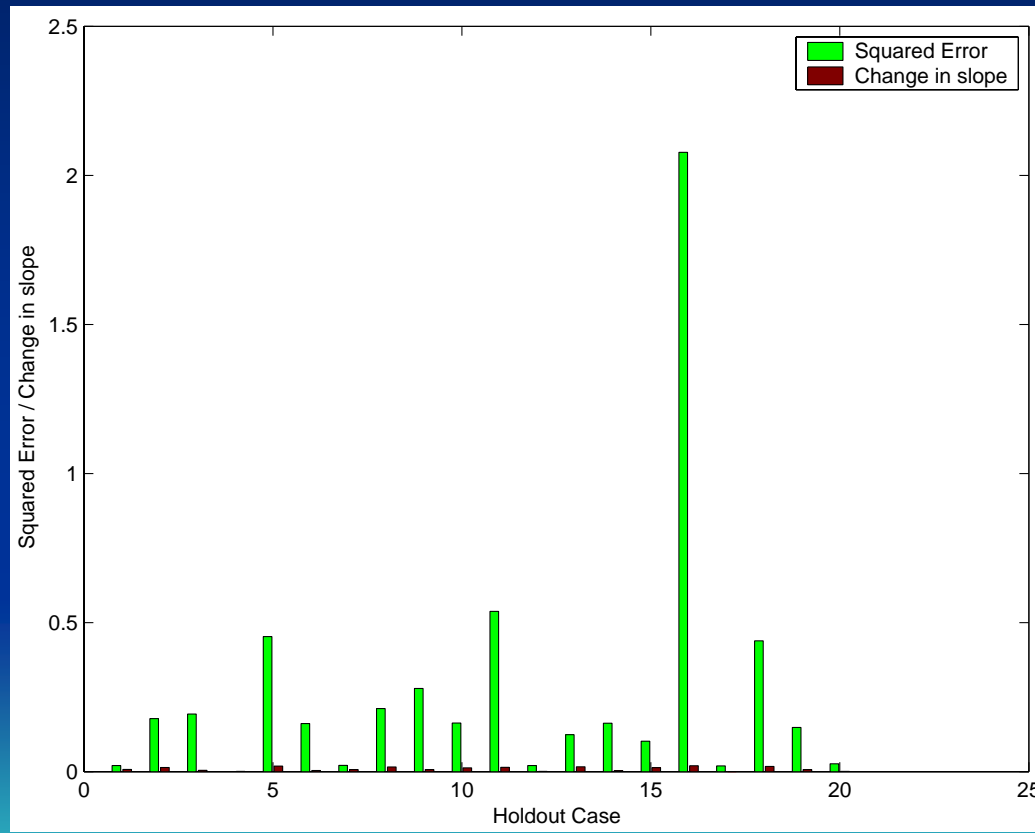
Comparisons

Optimal Linear Combination of PEAQ MOVs: Predictor Fit



Comparisons

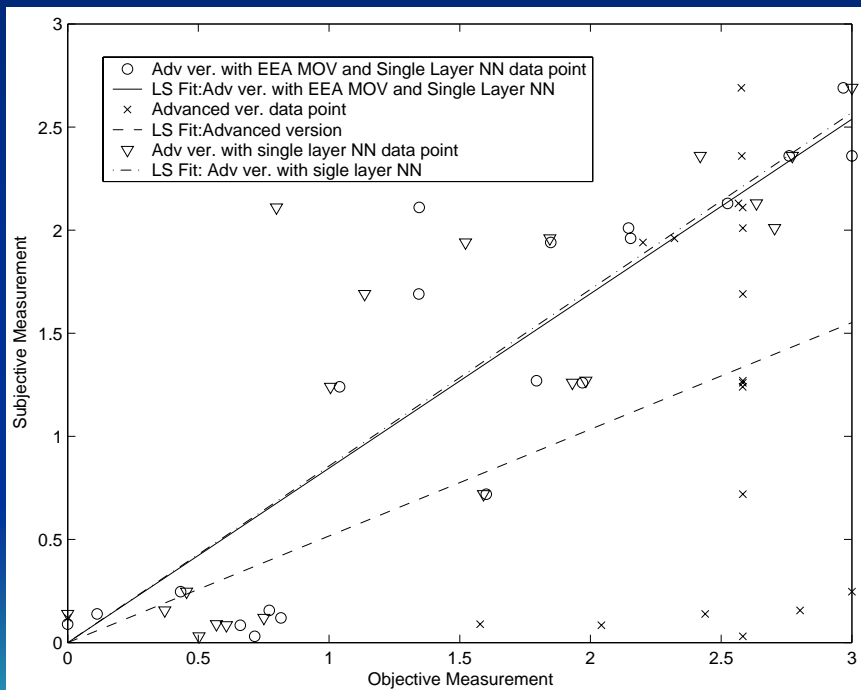
Optimal Linear Combination of PEAQ MOVs: Holdout Case



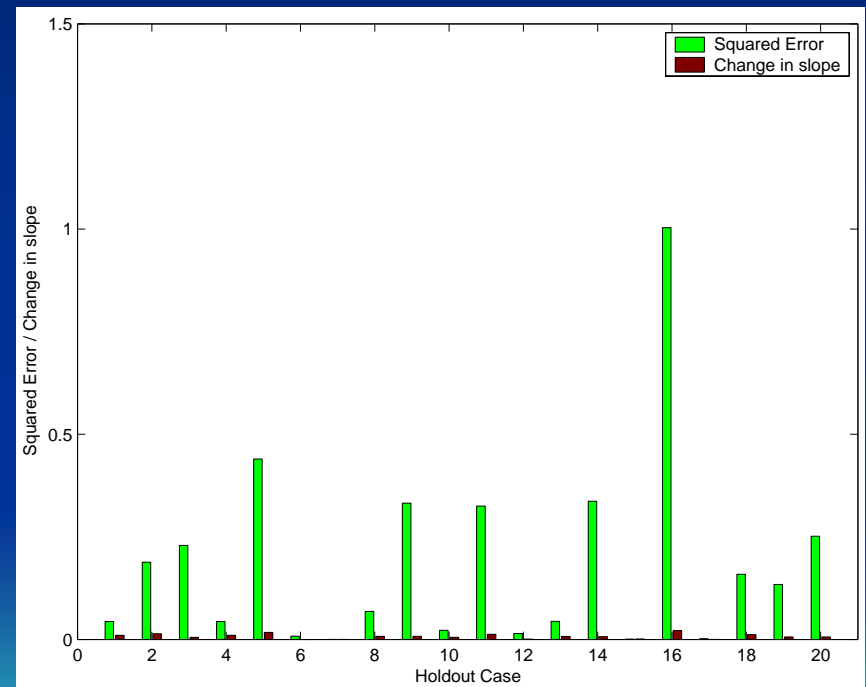
Comparisons

Optimal Linear Combination, PEAQ MOVs plus EEA

Optimal Fit



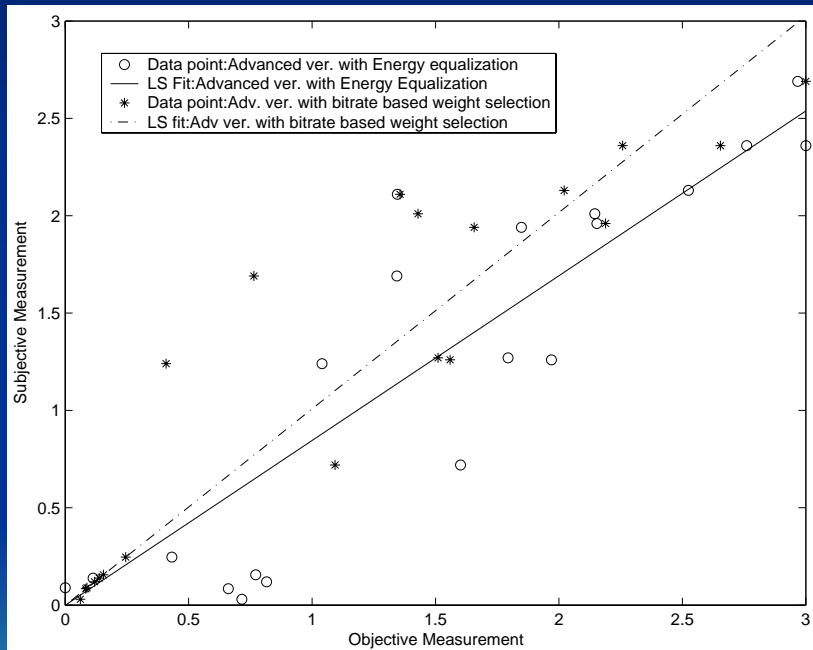
Error in Holdout Case



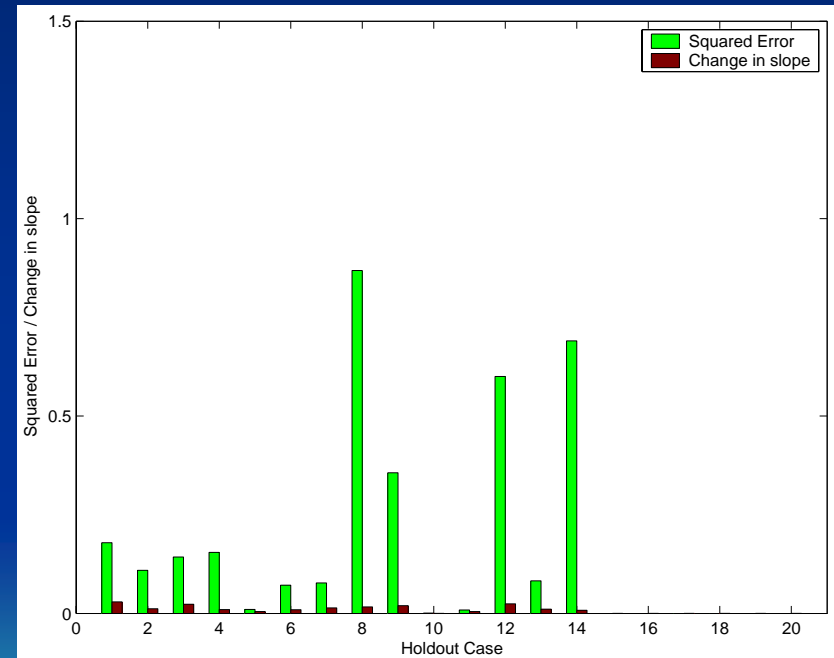
Comparisons

Optimal Linear Combination, Bitrate Optimized: Low/Mid Quality Audio

Optimal Fit



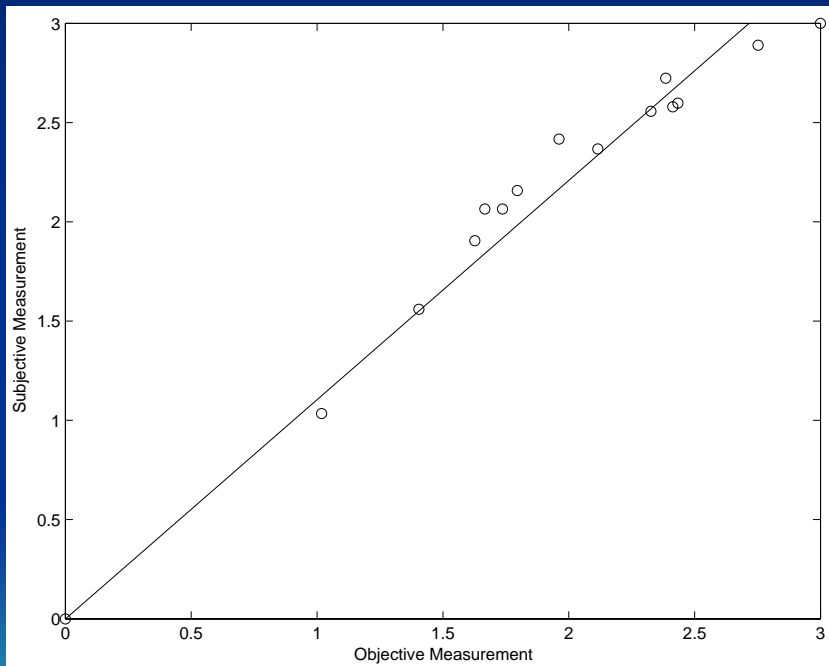
Error in Holdout Case



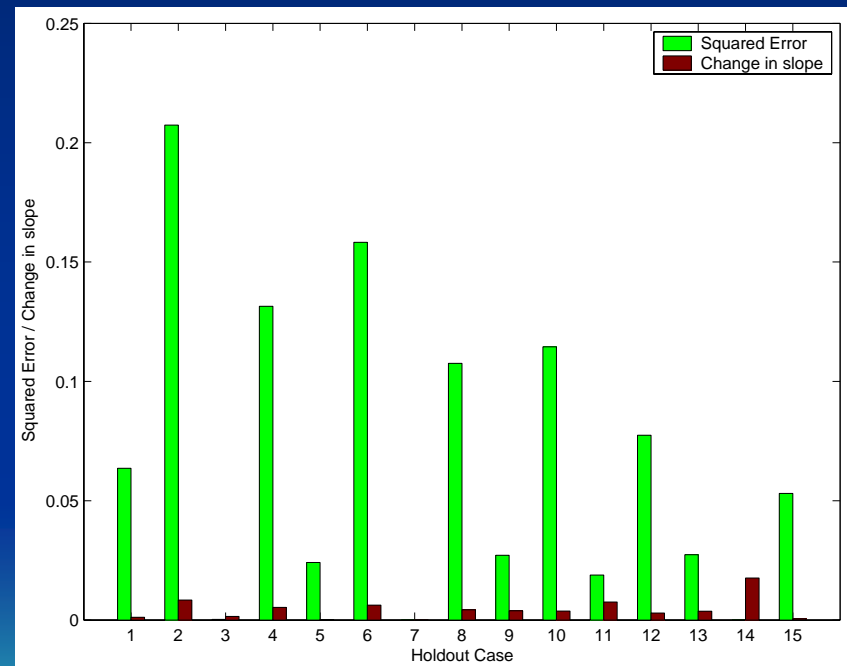
Comparisons

Optimal Linear Combination, Bitrate Optimized: High Quality Audio

Optimal Fit



Error in Holdout Case



Note: Perceptual measurements are simulated by treating the ODG values Generated by PEAQ-advanced as if they SDG values acquired through perceptual testing

Conclusions

- Combining the EEA truncation threshold with the PEAQ MOVs clearly improves the predictive performance of the metric
 - The correlation coefficient is increased
 - The MSE of the predication error is decreased
- If bitrate information is also available, performance is further increased significantly



Future Work

- Design a more complex 3-layer neural network similar to that used in PEAQ to generate the metric's output from the MOVs
- Generate additional subjective data using the more recent MUSHRA testing protocol and use it to more thoroughly validate the proposed metric

