INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

A Data-driven Cognitive Salience Model for Objective Perceptual Audio Quality Assessment

Pablo M. Delgado and Jürgen Herre

Paper/Poster N° 1857

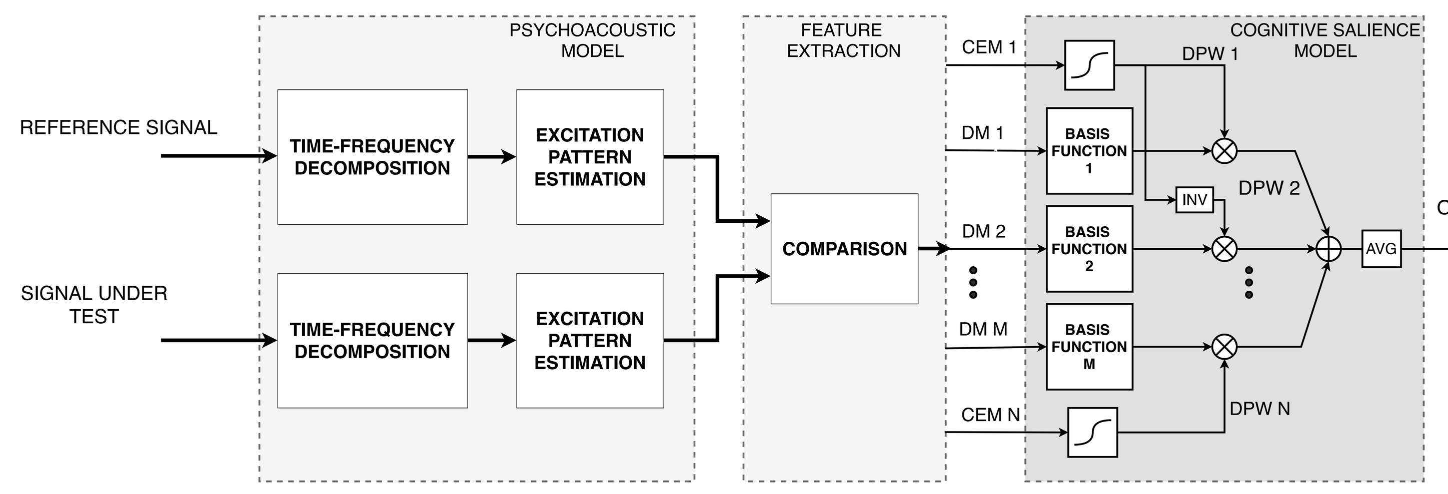
1. Introduction

- **Listening tests** are considered the **golden standard** for measuring audio codec quality, but they **take time and resources**.
- Objective quality assessment systems (OQAS) are **algorithms** that aim to **automatically predict Mean Opinion Scores (MOS) from listening tests**, to save time in development. Most systems **map different distortion metrics (DM) to an overall quality score** by using multivariate statistical learning algorithms with MOS as learning data [1].
- **A proposed Cognitive Salience Model (CSM)** uses **cognitive effect metrics (CEM)** to model **distortion salience** and **adaptively weight the importance of each DM** in the quality mapping stage.
- In terms of MOS prediction, our experiment shows that **OQAS with the CSM outperform systems that use other statistical learning methods** in the mapping stage for the same number of input variables.

2. Model Assumptions

- Certain cognitive effects in auditory perception **influence the salience of paritcular distortion types**. These effects influence the perceived severity of perceived degradations.
- The influence of these cognitive effects on distortion salience is dependent on the nature of the processed signal (stationary or transient-like, speech-like or music-like, etc...)
- A salient distortion will play a significant role in causing quality degradation of a given treated/processed signal, for a varfiety of treatments.

3. Proposed Approach



DM = Distortion Metric
CEM = Cognitive Effect Measure
DPW = Detection Probability Weight

- We use an *intrusive* OQAS based on the perceptual model of the **Perceptual Evaluation of Audio Quality (PEAQ – ITU-R BS-.1387-1)** method [1]. The DMs (M=6) correspond to the **Model Output Values (MOVs) of PEAQ's *advanced* version**. They describe degradations in terms of roughness, noisiness, dullness and other dimensions.
- The cognitive effects modelled through N=3 CEMs include a speech/music signal discriminator, and measurements for the amounts of *informational masking (IM)* and *perceptual streaming (PS)*. [2].
- The mapping stage of PEAQ is replaced by a CSM. Different CEM values weight the importance (salience) of certain DM values in the overall quality. To determine which CEMs significantly interact with a certain DM's salience, a salience metric is defined as a Pearson's correlation coefficient (CC):

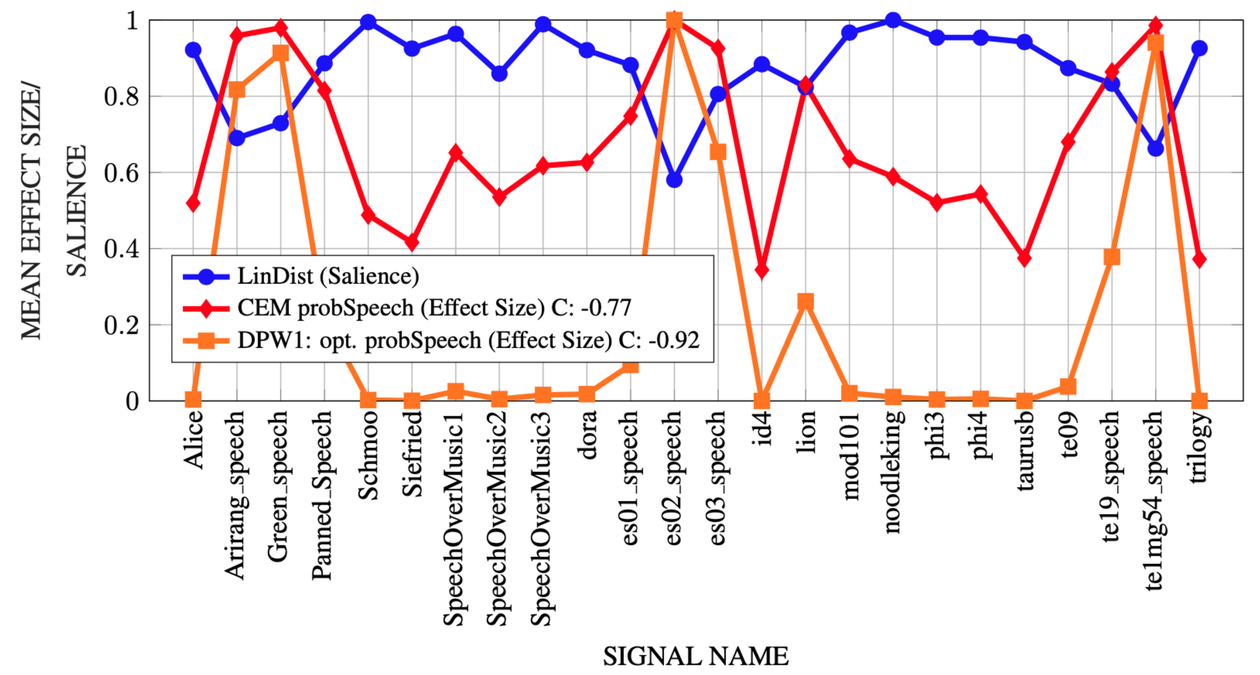
$$\mathcal{S}_m(j) = \frac{\sum_{i=1}^{I}(y_{ij}-\overline{y}_j)(BF_{mij}-\overline{BF}_{mj})}{\sqrt{\sum_{i=1}^{I}(y_{ij}-\overline{y}_j)^2}\sqrt{\sum_{i=1}^{I}(BF_{mij}-\overline{BF}_{mj})^2}}\,. \quad (1)$$

calculated over j input signals over all i treatments. The CC is calculated between the Basis Function outputs BF_{mj} that map DMs to quality, and the MOS y_ij in a calibration database.

- An additional interaction metric describing covariance between DM saliences and CEM values $C_m(\mathcal{S}_m, CEM_n)$ is used for the model parameter selection. Only DM_m, CEM_n pairs with strong interactions C_m will be used in the CSM.
- The values of C_m have been optimized using sigmoids as *Detection Probability Weights (DPW)* for modeling a cognitive effect's *detection probability* and *saturation* regions.

4. Performance Evaluation

- We evaluated the MOS prediction performance of an OQAS with CSM and other state-of-the-art OQAS in the USAC VT DB [3].



System	R	RMSE*
ViSQOL NSIM	0.82	5.6
PEAQ DI	0.69	8.1
DM + CEM	0.84	5.1
PROPOSED	0.86	4.6
PROPOSED (Opt.)	**0.90**	**3.7**

- We found significant interactions between speech/music discriminator output, and noise loudness and band limitation DMs. Also CEMs describing IM and PS predicted the salience of band limitation and Noise-to-Mask ratio DMs.
- The systems PROPOSED and PROPOSED (Opt.) **use the CSM and perform significantly better** than an OQAS (DM+CEM) that uses the same input variables, but maps the inputs to quality using an Artificial Neural Network as in [1]. The **use of DPW in the CSM of PROPOSED (Opt.) further increases performance**.

References

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

[2] Jayme Garcia Arnal Barbedo and Amauri Lopes, "A new cognitive model for objective assessment of audio quality," *J. Audio Eng. Soc*, vol. 53, no. 1/2, pp. 22–31, 2005.

[3] ISO/IEC JTC1/SC29/WG11, "USAC verification test report N12232," Tech. Rep., International Organisation for Standardisation, 2011.

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
Fraunhofer IIS