AUDIO LABS

AN IMPROVED METRIC OF INFORMATIONAL MASKING FOR PERCEPTUAL AUDIO QUALITY MEASUREMENT

Pablo M. Delgado
Jürgen Herre

Friedrich-Alexander-Universität Erlangen-Nürnberg

Fraunhofer IIS

Motivation
Impact of perceptual audio coding

Audio codec development is a very active branch of research,

The race for more **efficiency** and **best quality** continues as multimedia demand increases…

Crucial: A fast **time to market**.

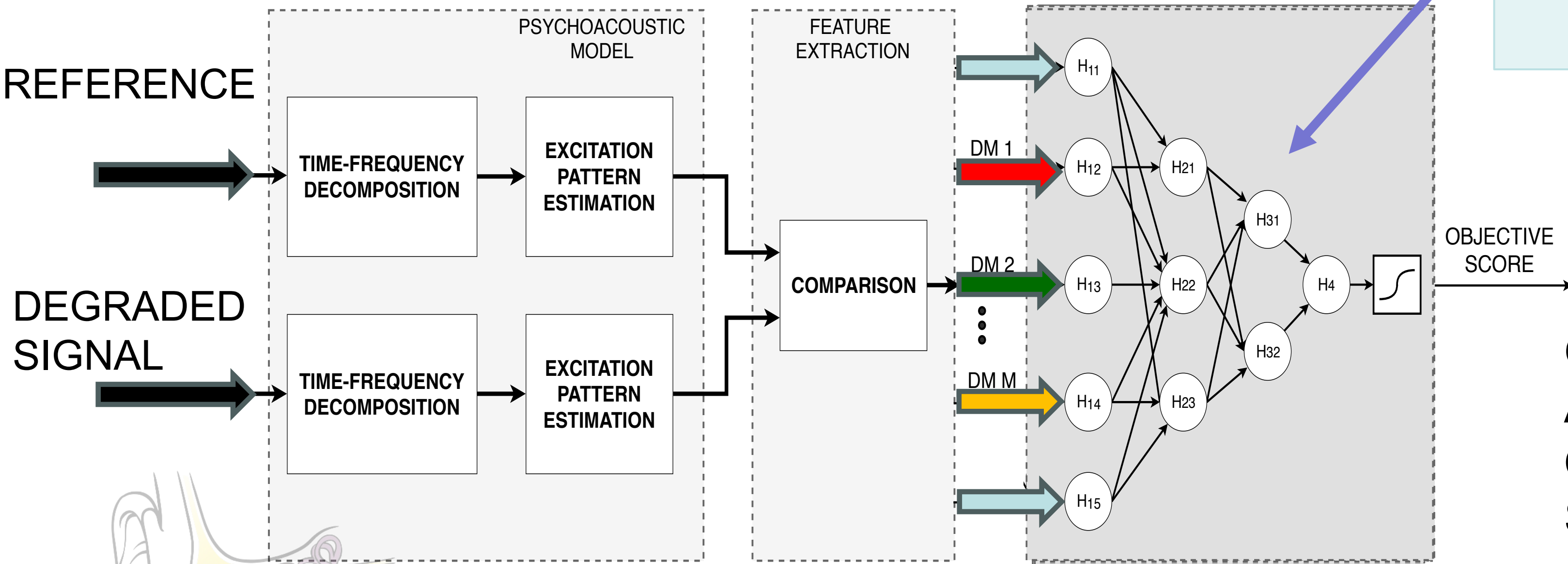
Subjective audio quality assessment can be cost-prohibitive,

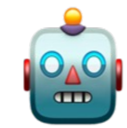
Objective audio quality assessment: algorithms that try to mimic subjective assessment, saving time and resources.

Objective Perceptual Quality Assessment

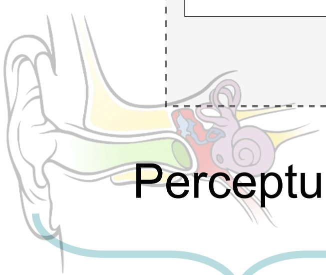
- Based on models of human auditory perception.
- Intrusive case (needs an undegraded reference signal)

Training Data (subjective scores)





Overall Audio Quality Score



Perceptual Model

Quality mapping
(e.g., **ANN**)

Loudness, energetic masking thresholds, roughness analysis, etc…

Distortion metrics: **Disturbance loudness**, Roughness, **Coloration**, **etc…**

Motivation

State of the art and challenges

- Many objective quality assessment methods have had great success with this perceptual approach:
  - Reference and coded signals are compared in the **perceptual domain,** instead of comparing waveforms**.**
  - Some, standardized for audio and speech coding (PEAQ, PESQ, …)

- However, newer coding technologies present a challenge:
  - Coded waveform is no longer preserved → **difficult** derivation of **perceptual representations.**
  - Tradeoff: **sensitivity** to small impairments versus **robustness** against parametric representations of the waveform.

Motivation

Degradation due to bandwidth extension (BWE)

- Bandwidth extension (parametric technique):
  - Saves bandwidth by **only transmitting the LF part of the spectrum** and **replicating the remaining HF spectrum at decoding time**
  - → improves quality at low target bitrates (~ 8-24 kbps), especially on **music signals**

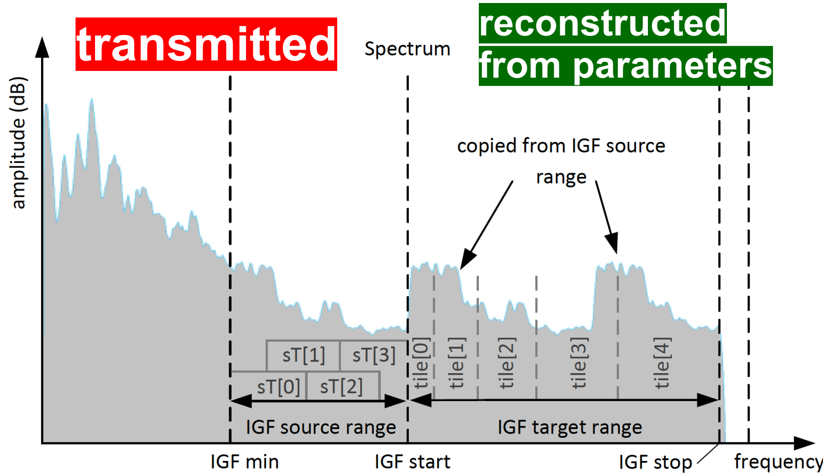
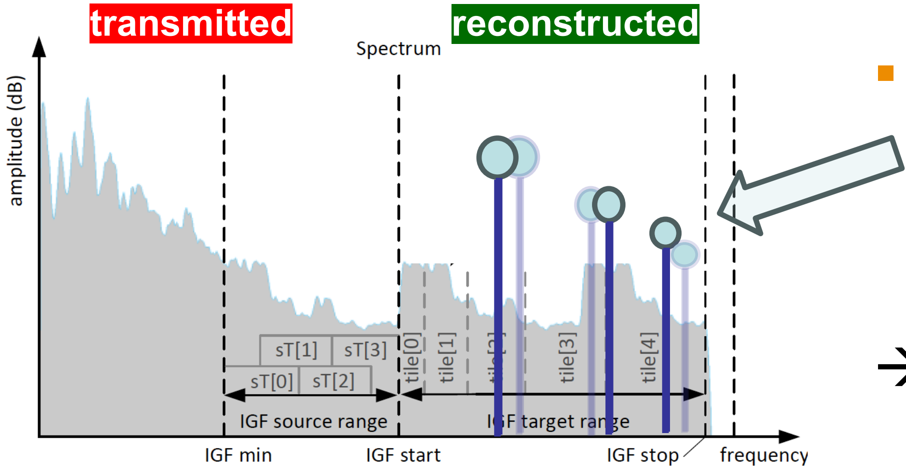


Fig. 5: IGF tile selection. [Disch, 18]

Motivation

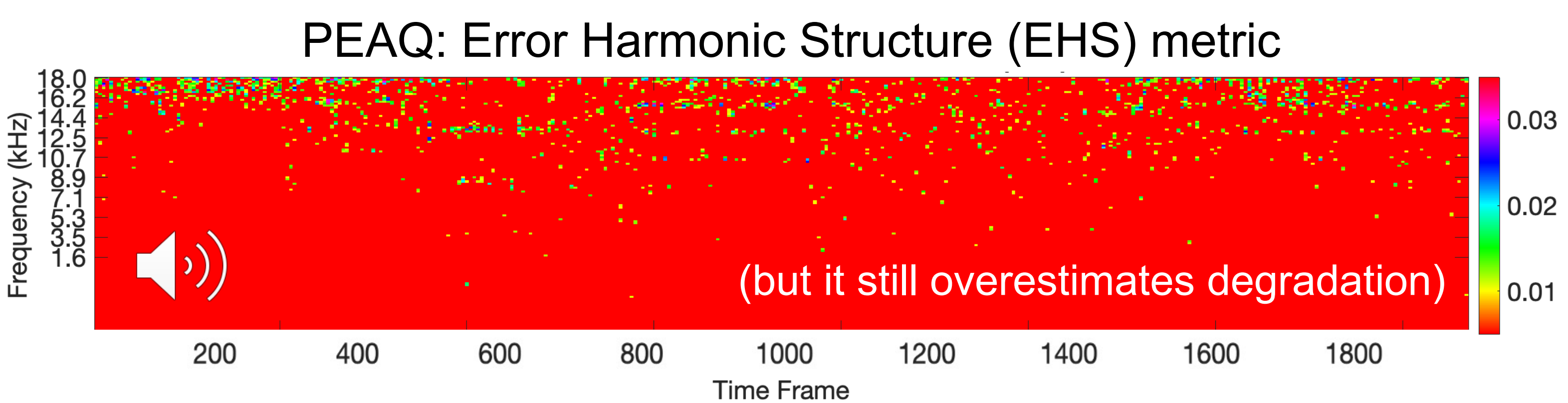
Degradation due to bandwidth extension (BWE)



- Errors in **position** and energy of **tonal** components, **rapidly varying** (i.e., modulation at frame rate)

→ roughness, tone beating, etc…

Mostly errors with **harmonic structures,** PEAQ measures them:



Proposed approach
Informational masking (IM) principle

- Not all harmonic errors are severe, or even audible. Can we predict their severity?
- **Informational masking (IM)** is a cognitive phenomenon that regulates the audibility/saliency of disturbances:

> **Increased** signal complexity (i.e., variability) **reduces** the audibility and perceived severity of disturbances.

- A **cognitive** effect metric (CEM) of **IM** can **interact** with **distortion** metrics to improve quality degradation prediction.[Beerends, 96]

Proposed approach
Previously used IM metrics

- Previous approach, calculate **total power deviation** :
  - $E_R(n,k)$: perceptual representation of reference signal at frame n and perceptual band k.
  - $\bar{p}(n,k)$: mean signal power over a 20ms time window. [Lutfi, 93]

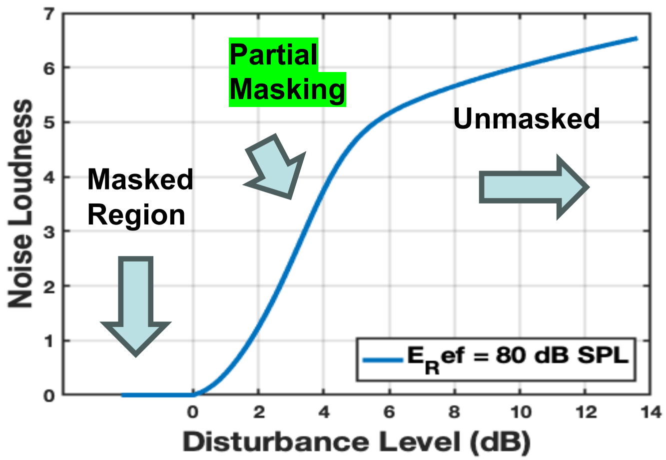
$$\mathcal{PDEV}(n,k) = |E_R(n,k) - \bar{p}(n,k)|$$

$$\mathcal{IM}(n) = \mathcal{PDEV}(n) = \frac{1}{K}\sum_{1}^{K}\mathcal{PDEV}(n,k)$$

- **But**: IM is stronger on **nearly-masked disturbances**, not for all disturbance strengths.

Proposed approach
Proposed IM metric

- Calculate signal variance **near the masking threshold**.
  - PEAQ's noise loudness metric describes behavior in this region:



$$N' = k\left[\left(1 + \frac{max(s_T E_T - s_R E_R, 0)}{E_{th} + s_R E_R \beta}\right)^{0.23} - 1\right]$$

$$\beta = e^{-\alpha(E_T - E_R)/\bar{E}_{Ref}}$$

Behavior in the partial masking region
[Thiede, 98]

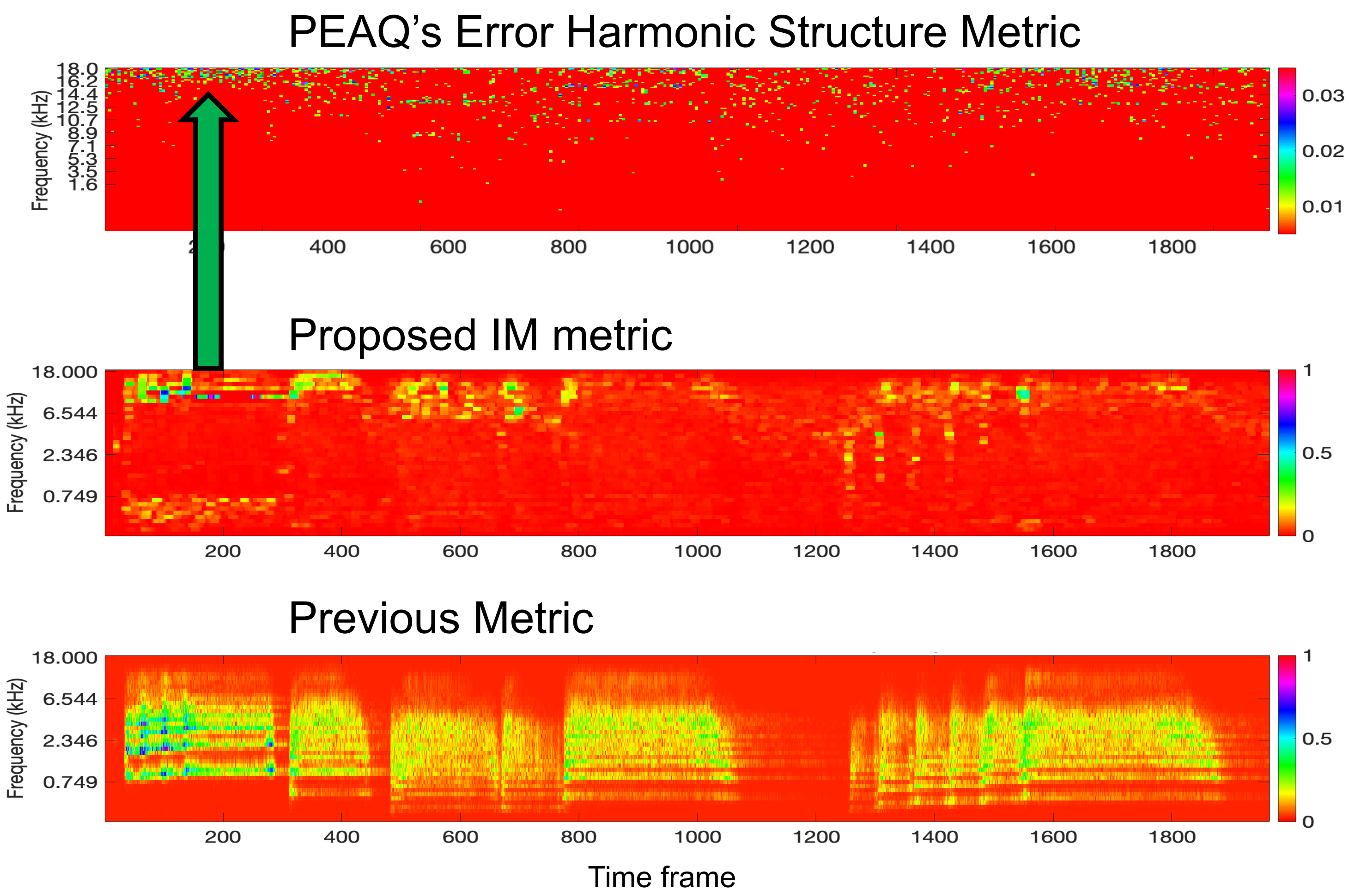
-

$$\mathcal{IM}(n) = \beta\text{-}\mathcal{VAR}(n) = \frac{1}{K}\sum_{1}^{K} var(\beta(n,k))$$

Proposed approach

Use IM to regulate severity of PEAQ's EHS metric

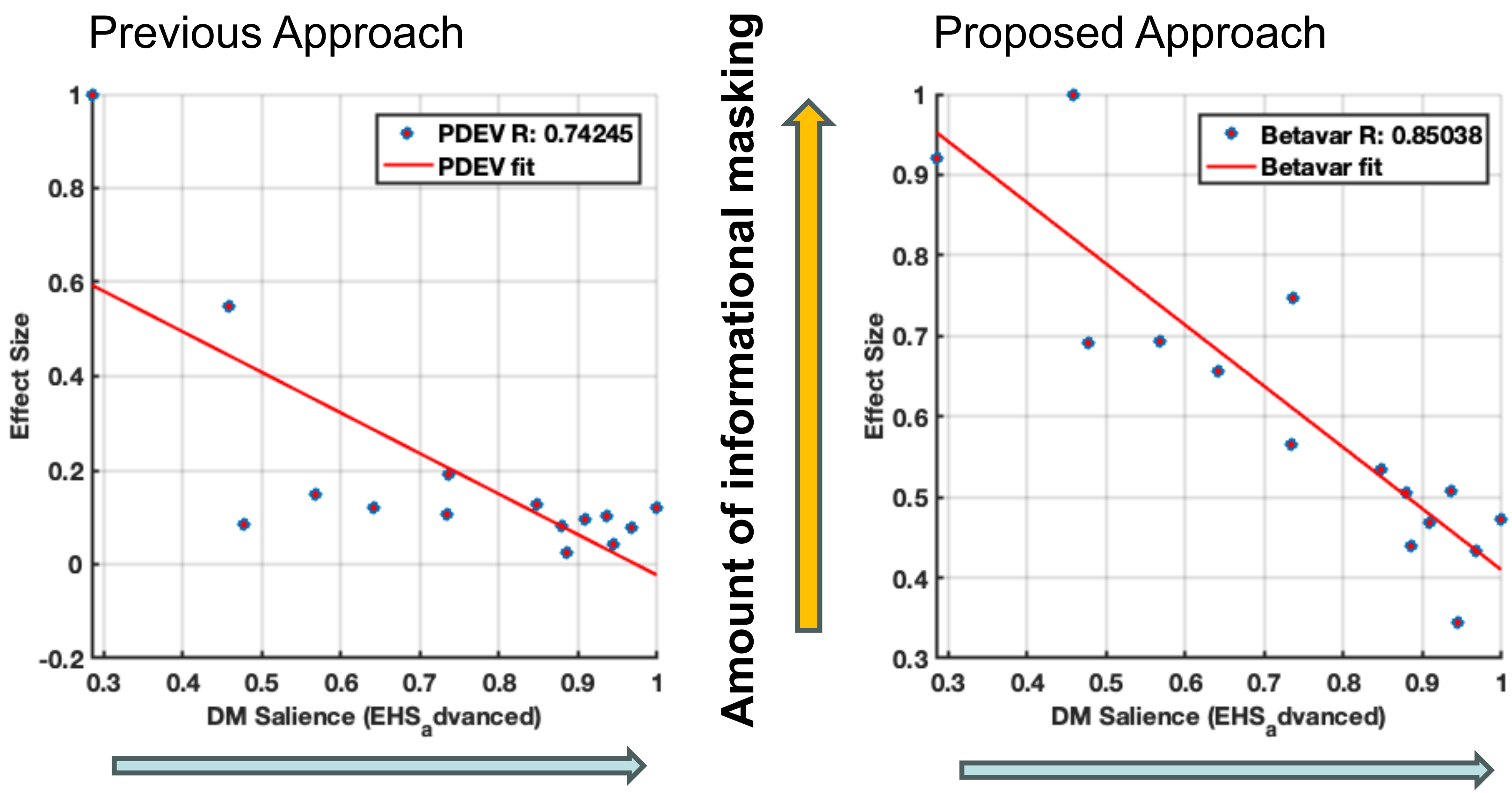
Regulates EHS severity at HF regions, where BWE errors happen



Does not regulate severity where BWE errors happen

Results

Proposed IM predicts EHS saliency

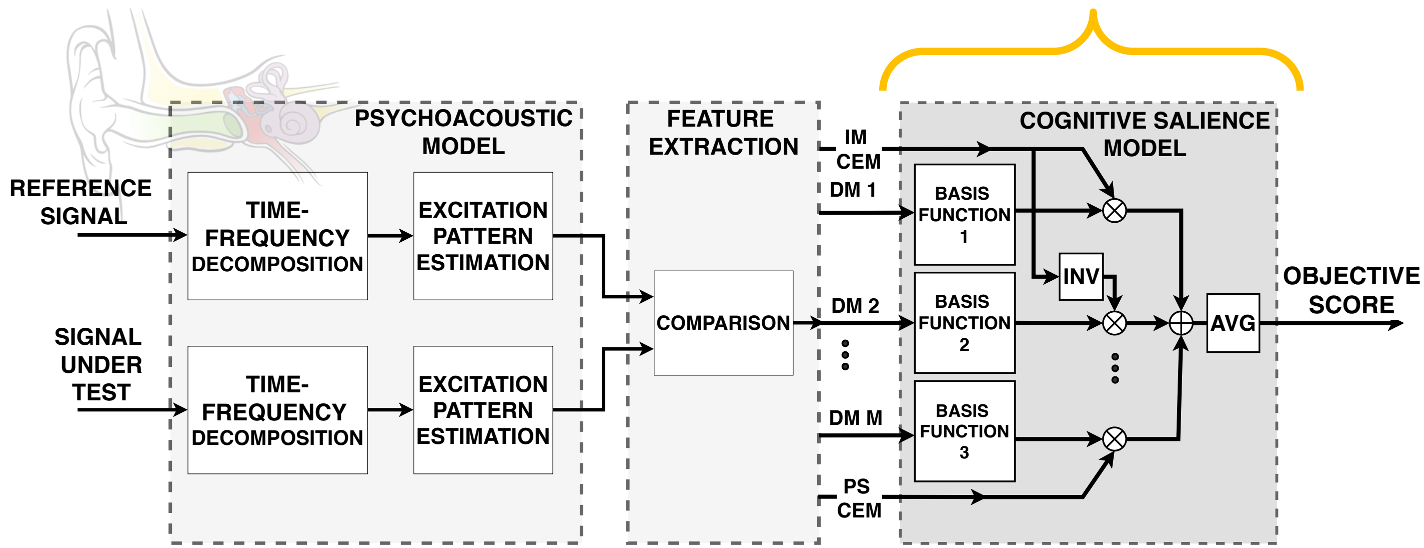


Harmonic error saliency→ usefulness of EHS **Harmonic error saliency→ usefulness of EHS**

AUDIO LABS

Incorporating IM into PEAQ (ITU-R BS.1387-1)

- Metric to quality mapping: **Cognitive Salience Model** (CSM)
  - Higher accuracy than other quality mapping methods [Delgado, 22]



IM = Informational Masking DM = Distortion Metric
PS = Perceptual Streaming CEM = Cognitive Effect Metric

- Distortion Metrics: PEAQ Advanced 5 MOVs.
- Cognitive Effects: IM, Perceptual Streaming. [Beerends, 96]
- Interactions modeled by a data-driven salience metric.

Validation
Subjective listening test databases

- 639 scores, 45 signals

	ELD A12	USAC VT 1	USAC VT 3
Reference	ISO/IEC JTC1/SC29/WG11 Verification of ELD (N10032)	ISO/IEC JTC1/SC29/WG11 Verification of USAC (N12232)	ISO/IEC JTC1/SC29/WG11 Verification of USAC (N12232)
Content	Speech, music, w/background noise and reverb, interfering speakers	Speech, Music, Mixed	Speech, Music, Mixed
Format	Stereo	Mono	Stereo
Type	MUSHRA	MUSHRA	MUSHRA
Mean quality	81.7	58.3	73.7
Codecs	AAC-ELD, **ELD-SBR**, AAC-LD, **HE-AAC**, G.722	**USAC**, AMR-WB+, **HE-AACv2**	**USAC**, AMR-WB+, **HE-AACv2**
B.Rate (kbps)	32-64	8-24	32-96
N	231	216	192

Results

Overall system performance

Correlation (R) between subjective scores and objective predictions

The system with the proposed IM model shows the **strongest R for all DBs**

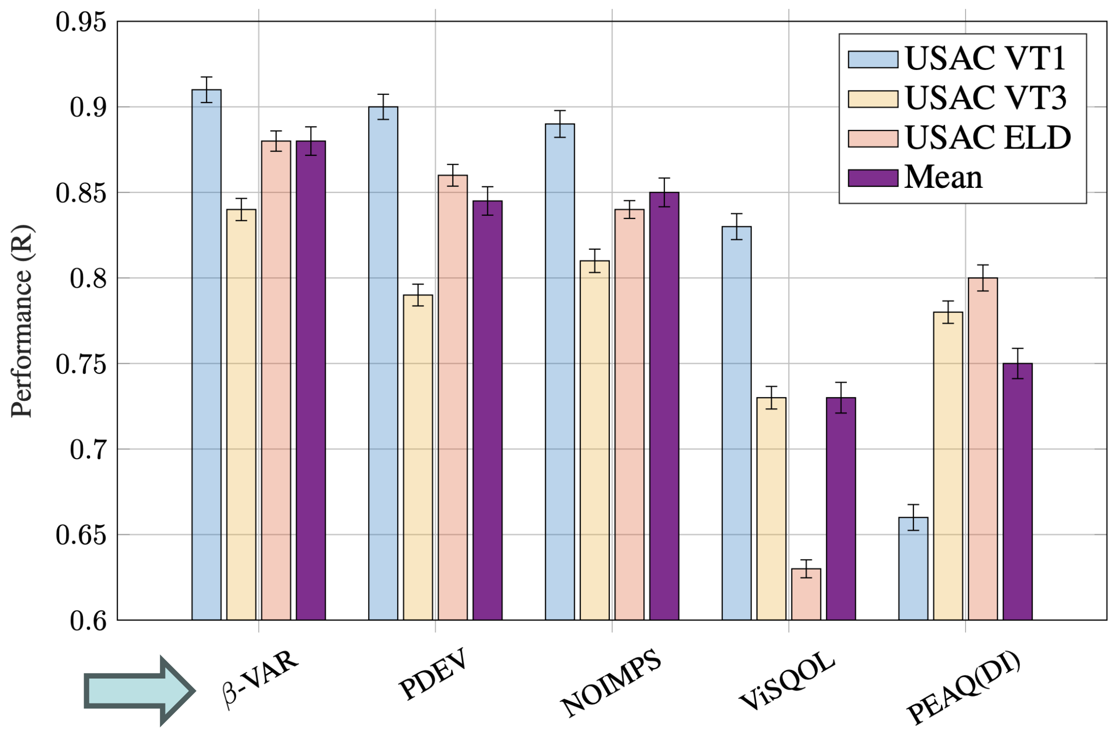


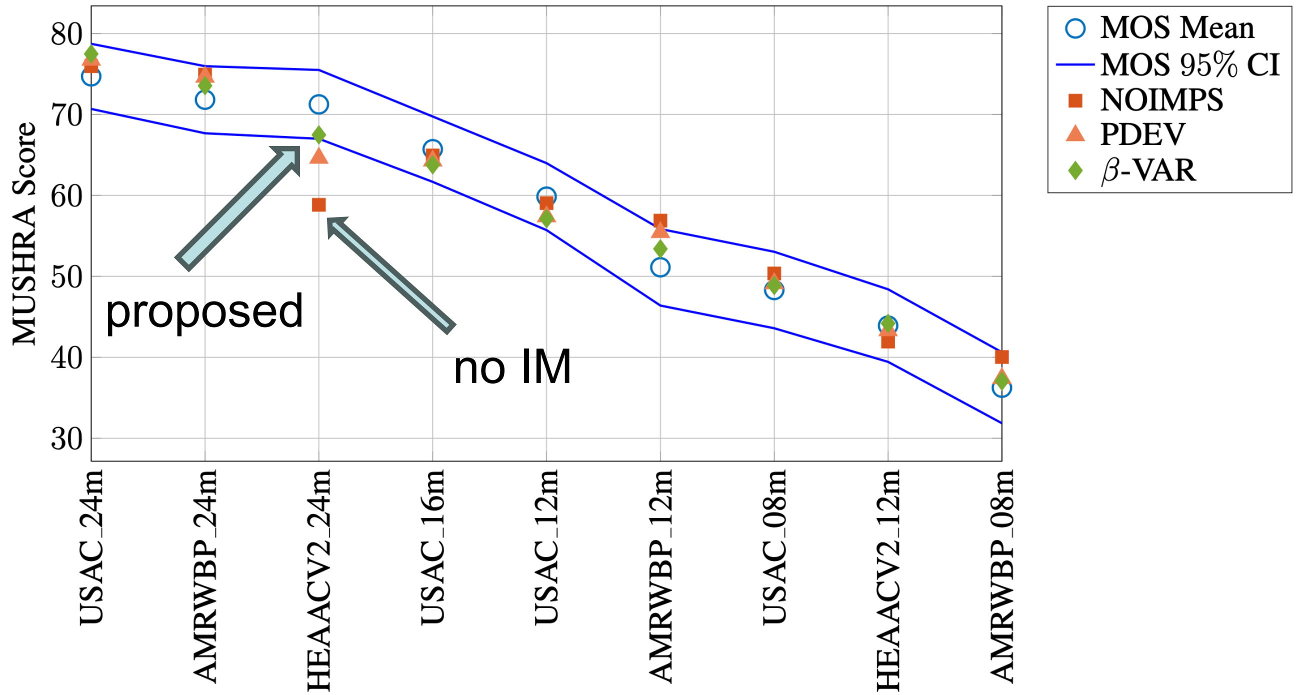
Figure 2: Overall system performance. Error bars denote 95% confidence intervals ($CI_{95\%}$).

Results

System performance for music items

- Music items coded with parametric techniques:

Using the proposed IM model
mitigates objective over-estimation
of quality degradation



Summary
Informational Masking in Quality Measurement

- Advanced audibility models enhance audio signal quality predictions by, e.g., incorporating cognitive effects.

- We presented a refined **informational masking (IM) model** considering **disturbances near the masking threshold**.

- The proposed IM metric notably improves predictions for **bandwidth-extended music signals** compared to other quality measurement systems.

Future Work

Interaction of cognitive models and distortion metrics

- Other cognitive models have been known to influence disturbance audibility and severity e.g., : co-modulation masking release (CMR), perceptual streaming, etc.

- Extend this concept to spatial sound quality perception.

THANK YOU!

QUESTIONS?

References

- ITU-R Rec. BS.1387 (2001). Method for objective measurements of perceived audio quality. Geneva,Switzerland.
- Beerends, J. G., van den Brink, W. A. C., and Rodger, B. (1996). The role of informational masking and perceptual streaming in the measurement of music codec quality. In Audio Engineering Society Convention 100, Copenhagen.
- Pablo M Delgado and Jürgen Herre, “A data-driven cognitive salience model for objective perceptual audio quality assessment,” in International Conference on Acoustics, Speech and Signal Processing (ICASSP).IEEE, 2022, pp. 986–990.
- Thiede, T. (1999). Perceptual Audio Quality Assessment Using a Non-Linear Filter Bank. PhD thesis, Fachbereich Elektrotechnik, Technische Universität Berlin.
- Robert A Lutfi, “A model of auditory pattern analysis based on component-relative-entropy,” The Journal of the Acoustical Society of America, vol. 94, no. 2, pp. 748–758, 1993.
- Disch, S., van de Par, S., Niedermeier, A., Burdiel Pérez, E., Berasategui Ceberio, A., & Edler, B. (2018, October). Improved psychoacoustic model for efficient perceptual audio codecs. In Audio Engineering Society Convention 145. Audio Engineering Society.