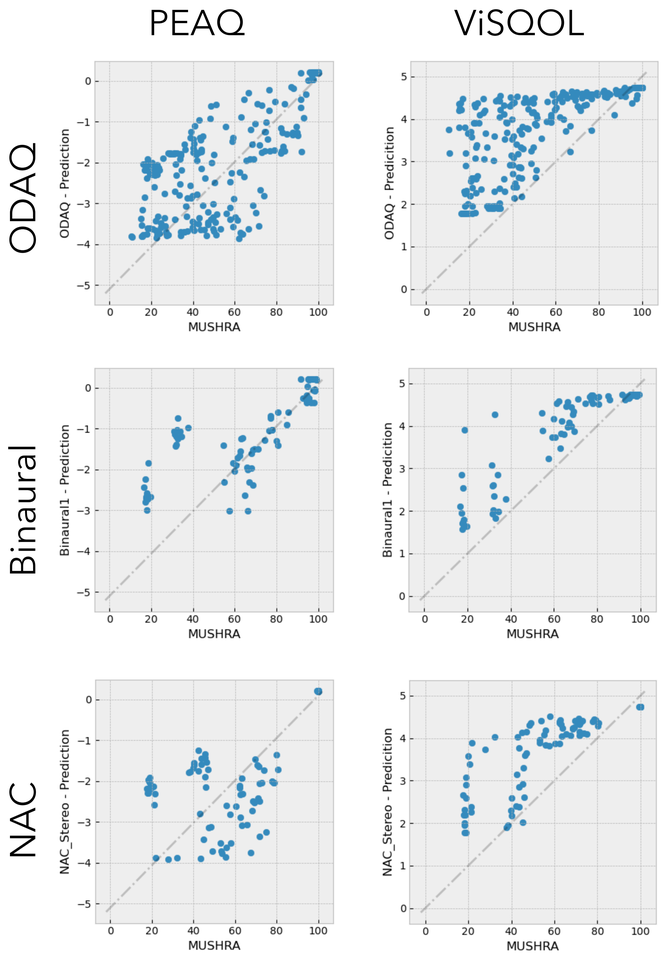
ENHANCED GENERATIVE MACHINE LISTENER

Vishnu Raj[1], Gouthaman KV[1], Shiv Gehlot[1], Lars Villemoes[2], Arijit Biswas[3]
[1] Dolby Technology India Pvt. Ltd., Bangalore, India.
[2] Dolby Sweden AB, Gävlegatan 12A, 113 30 Stockholm, Sweden
[3] Dolby Germany GmbH, Marienbergstr. 98, 90411 Nürnberg Germany.

THE PROBLEM

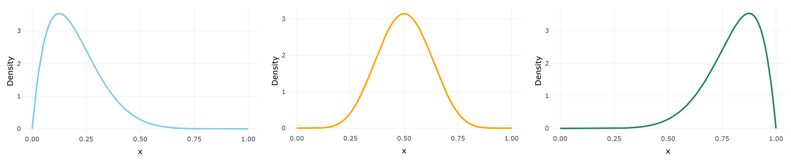
Existing metrics like PEAQ and ViSQOL often fail to correlate with human perception in complex audio signals and modern neural audio codecs (NAC).



GML WITH BETA LOSS

Beta distribution

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad x \in [0,1], \alpha > 0, \beta > 0$$



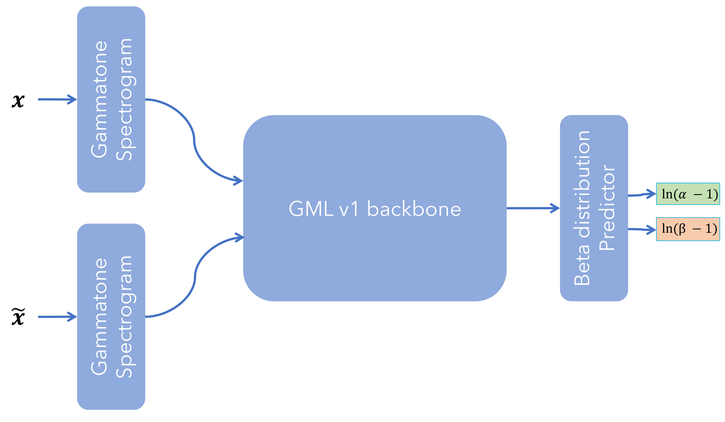
Low quality signals, such as anchor signals, score low and can be captured by Beta distribution with $\alpha < \beta$.

Mid quality signals, such as codec artifacts, can be captured by Beta distribution with $\alpha \approx \beta$.

High quality signals, such as reference-like, score high and can be captured by Beta distribution with $\alpha > \beta$.

$$\mathcal{L}(x, \tilde{x}, s) = -(\alpha - 1)\ \ln s\ -\ (\beta - 1)\ \ln(1 - s)\ +\ \ln B(\alpha, \beta)$$

Architecture

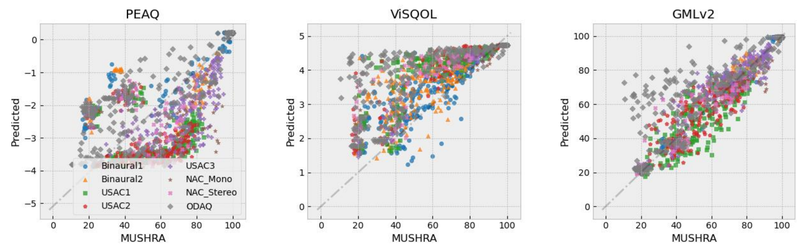


More human test (MUSHRA) data

- ✓ Added new 14688 MUSHRA datapoints on Neural Audio Codecs
- ✓ 82191 total MUSHRA datapoints

RESULTS

Results overview



Quantitative comparison

	PEAQ		ViSQOL		GMLv1		Proposed GMLv2	
Eval. set	$R_p^\uparrow$	$R_s^\uparrow$	$R_p^\uparrow$	$R_s^\uparrow$	$R_p^\uparrow$	$R_s^\uparrow$	$R_p^\uparrow$	$R_s^\uparrow$
USAC-1	0.4700	0.4000	0.8090	0.8356	0.8788	0.8779	**0.9191**	0.8964
USAC-2	0.4200	0.2000	0.7712	0.7775	0.8905	0.8574	**0.9261**	**0.8919**
USAC-3	0.5600	0.6200	0.8229	0.9036	0.9233	0.9443	**0.9442**	**0.9348**
Binaural 1	0.7500	0.7900	0.8976	0.9271	0.9827	**0.9544**	**0.9840**	0.9418
Binaural 2	0.4200	0.5600	0.9613	0.8539	0.9830	0.8968	**0.9895**	**0.9068**
NAC Mono	0.3400	0.3100	0.8920	0.8614	0.9514	0.9210	**0.9706**	**0.9398**
NAC Stereo	0.5800	0.4000	0.8157	0.8948	0.8689	0.8175	**0.9492**	**0.9254**
ODAQ	0.7100	0.6500	0.6995	0.7954	0.7092	0.7537	**0.8256**	**0.8287**
Agg. Score	0.5606	0.5187	0.8468	0.8643	0.9234	0.9006	**0.9526** ▲0.0292	**0.9205** ▲0.0199

Key takeaways

- ➢ Introduced **new Beta loss function**, which captures the human test MUSHRA scores better.
- ➢ GMLv2 **aligns better with human perception** on multiple datasets when compared to PEAQ and ViSQOL

References

1. Jiang, Guanxin, Lars Villemoes, and Arijit Biswas. "Generative machine listener." *AES Convention 155, Audio Engineering Society 2023.*
2. Biswas, Arijit, and Guanxin Jiang. "RF-GML: reference-free generative machine listener." *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025.

SCAN ME