

# 2025 LRAC Challenge

*Challenge overview including key findings*

**Organizers:** Wojcicki, Isik, Lechler, Balic, Yesilbursa, Zhang, Mack, Łaganowski (Cisco Systems)

**Co-Organizers and Academic Partners:** Watanabe (CMU), Kim (UIUC), Adi (HUJI)

May 4, 2026



# Challenge introduction

Kamil Wojcicki, *Cisco Systems*



# Why a challenge?

*Neural audio coding an active area of research, but:*

## Fragmented progress

- Different datasets, bitrates, metrics
- Limited noise and reverb eval
- Difficult to compare

## Controlled comparison

- Same joint constraints on bitrate, compute and latency
- Listening test materials and design

*The LRAC Challenge:*

## Coding + enhancement

- Dedicated track for systems that do both (e.g., joint or modular)
- Advantages for joint in low-resource

## Listener evaluation at scale

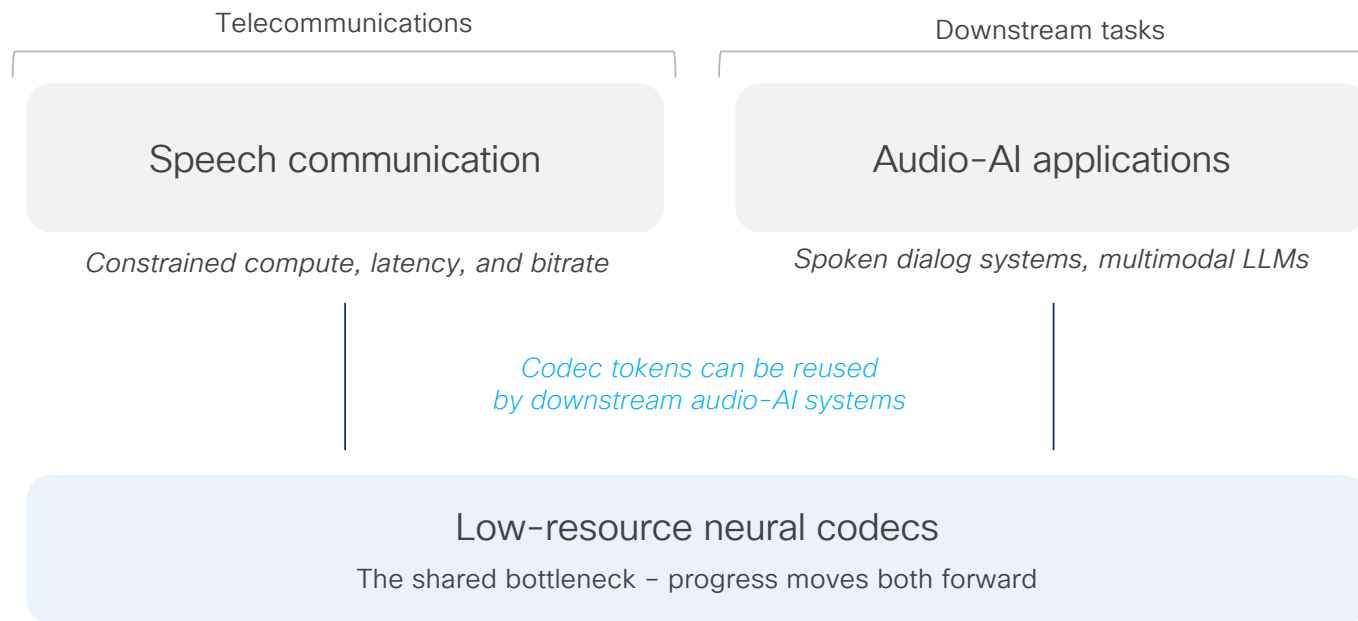
- Real-world noisy and reverberant
- Assessment of both quality and intelligibility
- Gather insights re: objective metrics

*LRAC delivers structured, comparable evaluation*



# Impact beyond coding

*Efficient speech representations are becoming foundational*



# Challenge Tracks

	Track 1 – Transparency	Track 2 – Enhancement
<b>Goal</b>	Speech coding: compress + preserve quality	Speech coding + denoise + dereverb
<b>Latency</b>	$\leq 30$ ms	$\leq 50$ ms
<b>Compute (total)</b>	$\leq 700$ MFLOP/s	$\leq 2600$ MFLOP/s
<b>Compute (receive)</b>	$\leq 300$ MFLOP/s	$\leq 600$ MFLOP/s
<b>Bitrate</b>	1 kbps + 6 kbps (both required, single decoder)	1 kbps + 6 kbps (both required, single decoder)

## Additional constraints

- 24 kHz mono · constant bitrate only
- Predefined training datasets: 702h speech + 335h noise + RIRs
- No external speech or noise data · no test-time adaptation



# Evaluation design and statistics

LBR: 6 kbps  
ULBR: 1 kbps



## Test phase – crowdsourced listening test battery

	Conditions	# Files	Max. Bitrates	Goal	Test	Test Range	Weight (ULBR / LBR)
<b>Track 1</b>	1a. Clean speech	90	1, 6	Quality on clean speech	MUSHRA-1S	[0, 100]	20% / 20%
	1b. Real-world light noise & reverb	90	1, 6	Robustness to light noise & reverb	DCR	[1, 5]	20% / 20%
	1c. Simultaneous talkers	20	1, 6	Preservation of multiple speakers	DCR	[1, 5]	5% / 5%
	1d. Speech intelligibility in clean	384	1	Speech intelligibility (English)	DRT	[-100, 100]	10% / -
<b>Track 2</b>	2a. Clean speech	90	1, 6	Quality on clean speech	MUSHRA-1S	[0, 100]	10% / 15%
	2b. Real-world reverb	60	1, 6	Dereverberation performance	ACR	[1, 5]	10% / 20%
	2c. Real-world speech in noise	100	1, 6	Denosing performance	ACR	[1, 5]	10% / 20%
	2d. Speech intelligibility in clean	384	1	Speech intelligibility (English)	DRT	[-100, 100]	5% / -
	2e. Speech intelligibility in noise	384	6	Speech intelligibility (English)	DRT	[-100, 100]	- / 10%

## Crowdsourced evaluation statistics

- 65 listening tests conducted
- 17 codecs evaluated at two bitrate modes on specialized test sets
- Approximately 186,720 individual ratings of audio files gathered
- 24,140 crowdsourced participants (not all unique) contributed to the evaluation



# Further details and supporting publications

- Challenge website: <https://lrac.short.gy>
- MUSHRA-1S: A scalable and sensitive test approach for evaluating top-tier speech processing systems (*Lechler & Balic*)  
[ICASSP 2026 / arXiv:2509.19219](#)
- Assessing speech quality metrics for evaluation of neural audio codecs under clean speech conditions (*Mack, et al.*)  
[ICASSP 2026 / arXiv:2509.24457](#)
- Low-Resource Audio Codec (LRAC): 2025 Challenge Description (*Wojcicki, et al.*)  
[2026 LRAC Workshop / arXiv:2510.23312](#)



# Track 1 Results

LBR: 6 kbps  
ULBR: 1 kbps

System	Clean speech		Real-world mild noise + reverb		Simultaneous talkers		Intelligibility in clean	Aggregate Score	
	ULBR	LBR	ULBR	LBR	ULBR	LBR	ULBR	Final Score	Rank
<b>Scale</b>	MUSHRA [0,100]		DMOS [1,5]		DMOS [1,5]		DRT [-100, 100]	Weighted Sum [0, 100]	
<b>Weight</b>	20%	20%	20%	20%	5%	5%	10%		100%
teamwzqaq	<u>62.65</u>	<u>81.75</u>	3.02	<b>4.44</b>	<b>2.82</b>	<b>4.35</b>	<u>85.43</u>	<b>71.91</b>	1
nano-codec	59.23	81.17	<u>3.13</u>	<b>4.44</b>	<u>2.60</u>	4.22	78.12	<u>70.86</u>	2
aitd-go	60.90	80.69	<b>3.40</b>	4.16	2.08	2.98	<b>85.57</b>	69.22	3
nju-aalab	<b>65.20</b>	<b>89.19</b>	2.74	4.12	1.70	2.82	82.98	67.48	4
boya-audio	35.22	77.24	2.21	<u>4.30</u>	2.03	<u>4.26</u>	80.29	59.42	5
pdura7	42.75	62.56	2.30	3.29	1.68	2.05	75.34	49.94	6
lrac-challenge	17.92	74.28	1.31	3.35	1.26	2.20	75.90	42.36	7

@Workshop

IRIS Codec  
NanoCodec  
PhoenixCodec  
Vocodec

Baseline

- 6 submissions + challenge baseline
- Top 3 systems within 2.7 points of each other
- All external entries beat the baseline by 8-30 points
- Winner: teamwzqaq (IRIS codec) – presents in next oral session



# Track 2 Results

LBR: 6 kbps  
ULBR: 1 kbps



System	Clean speech		Real-world speech in noise		Real-world speech reverb		Intelligibility in clean	Intelligibility in noise	Aggregate Score		
	ULBR	LBR	ULBR	LBR	ULBR	LBR	ULBR	LBR	Final Score	Rank	
Scale	MUSHRA [0,100]		MOS [1,5]		MOS [1,5]		DRT [-100, 100]	DRT [-100, 100]	Weighted Sum [0, 100]		
Weight	10%	15%	10%	20%	10%	20%	5%	10%		100%	
PR-Vocoder	nju-aalab	<b>67.30</b>	<b>87.47</b>	<b>2.66</b>	<b>3.42</b>	<b>3.04</b>	<u>3.80</u>	80.61	<u>72.09</u>	<b>68.32</b>	1
KD-Vocoder	xuyang	<u>63.66</u>	<u>87.10</u>	2.33	3.18	2.57	3.57	<b>84.46</b>	72.01	<u>63.64</u>	2
NanoCodec	nano-codec	56.32	85.76	2.16	3.08	<u>2.74</u>	<b>3.85</b>	67.69	68.34	63.01	3
	aitd-go	63.41	82.07	<u>2.48</u>	<u>3.37</u>	2.39	3.36	78.72	<b>73.47</b>	62.62	4
IRIS Codec	teamwzqaq	60.47	81.09	1.94	2.86	2.43	3.45	<u>81.78</u>	63.78	58.42	5
	boya-audio	49.74	78.17	1.92	2.91	2.38	3.66	<u>70.43</u>	70.39	58.08	6
Vocoder	leyan	58.63	75.96	2.09	2.89	2.15	3.05	76.21	66.35	55.28	7
	parslog	41.34	71.40	1.90	2.57	2.06	2.76	66.95	50.68	48.10	8
	pdura7	38.86	61.72	1.80	2.36	2.12	2.87	73.90	67.19	46.80	9
Baseline	lrac-challenge	21.26	60.06	1.30	2.25	1.39	2.32	75.86	68.21	38.52	10

- Total: 9 submissions + baseline
- nju-aalab wins by ~5 point margin
- Winner: nju-aalab (PR-Vocoder) – presents in next oral session

@Workshop

# Insights from the challenge

Yusuf Isik, *Cisco Systems*



# Comparison with high-resource baselines

Codec	Bitrate (kbps)	Params (M)	Compute (GFLOP/s)	Latency (ms)
DAC	6	74.7	120.0	220
Mimi	1.1	79.3	10.7	80
DualCodec	1.23	684	75.0	4000/offl.
SNAC	0.98	19.8	14.5	480
Teamwzqaq	1/6	4.0	0.7	30

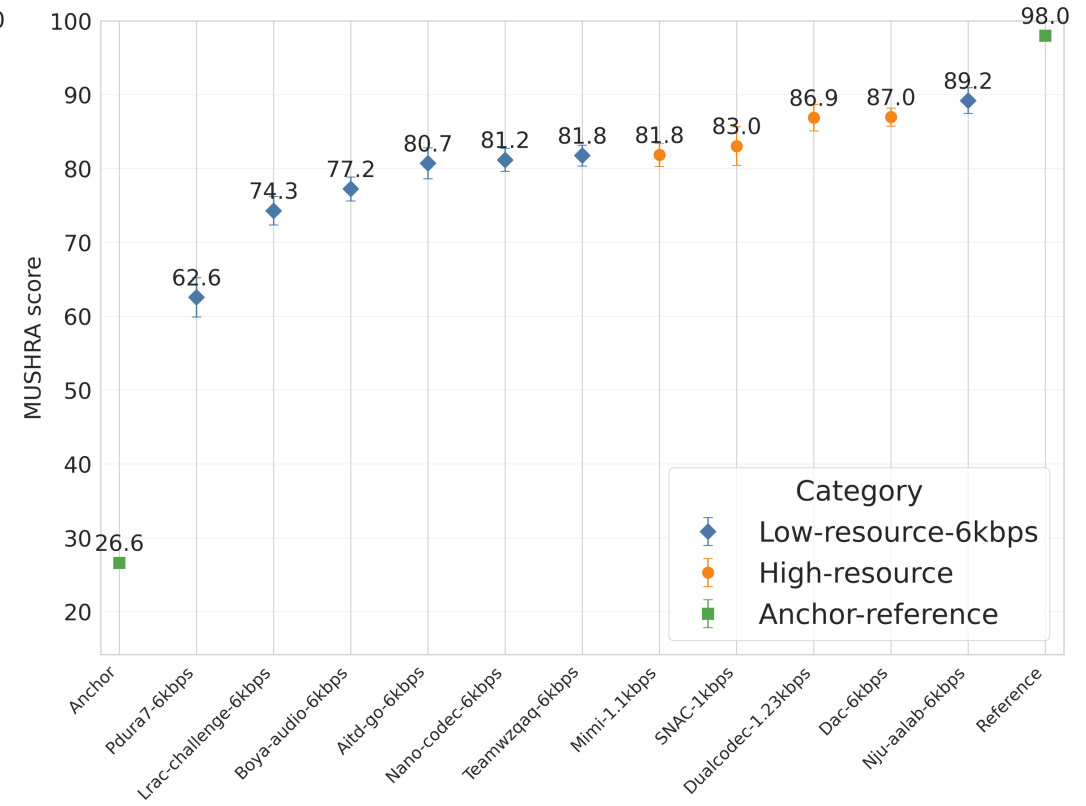
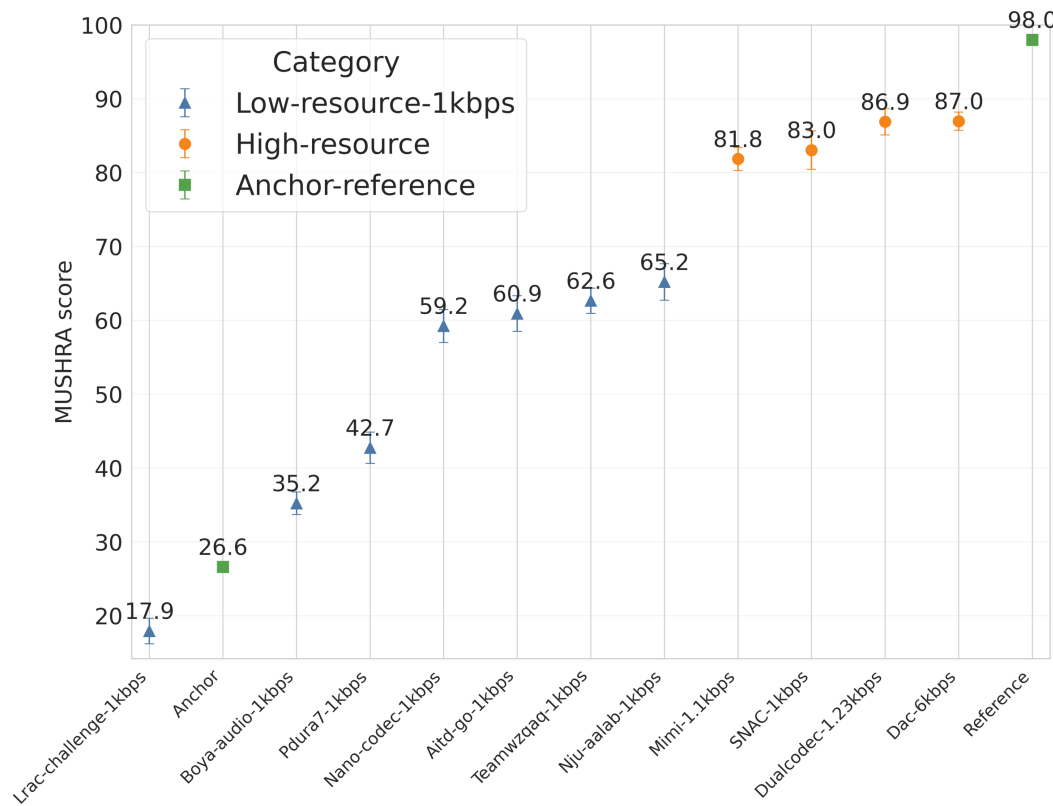
- Most SOTA high-resource codecs are one to two order of magnitudes more expensive than low-resource codecs in Track 1 of the LRAC Challenge.
- Criteria for selecting the SOTA high-resource codecs:
  - Has a variant working at 24 kHz sampling rate
  - Supports either ~1 kbps or ~ 6 kbps
  - DualCodec uses W2v Bert 2.0 working offline in addition to encoder. 4 sec latency without it, offline with the front-end



# Track 1: Clean speech MUSHRA test

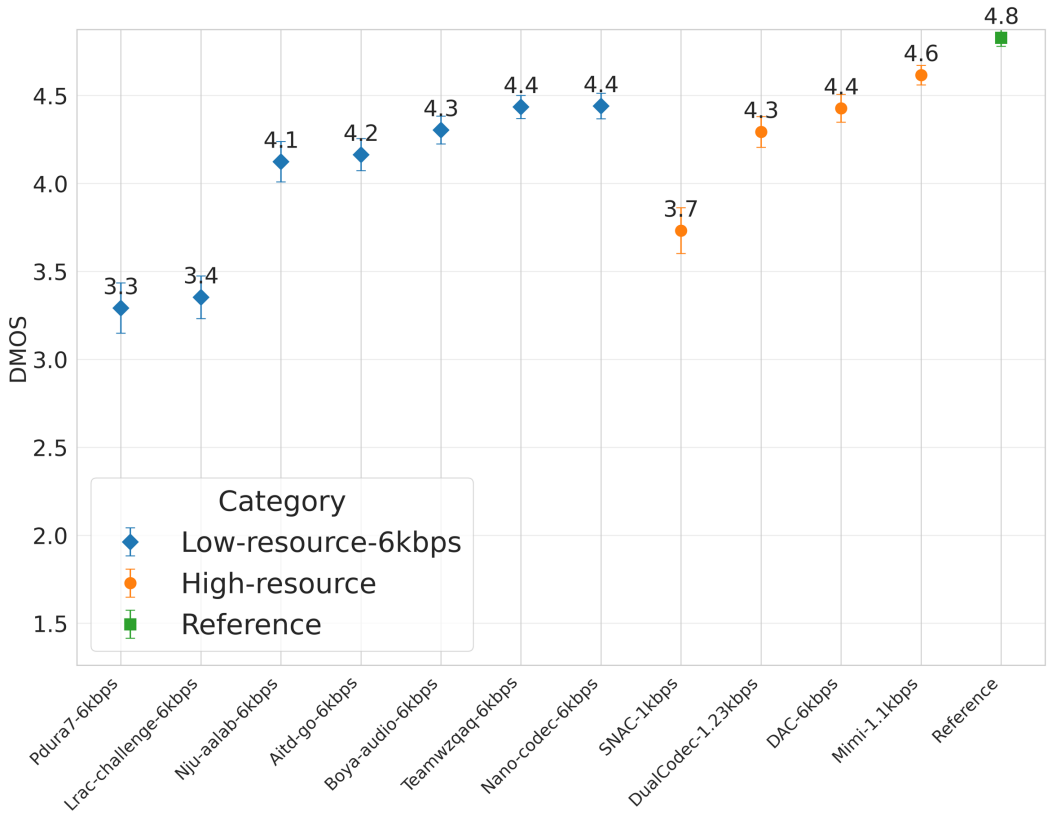
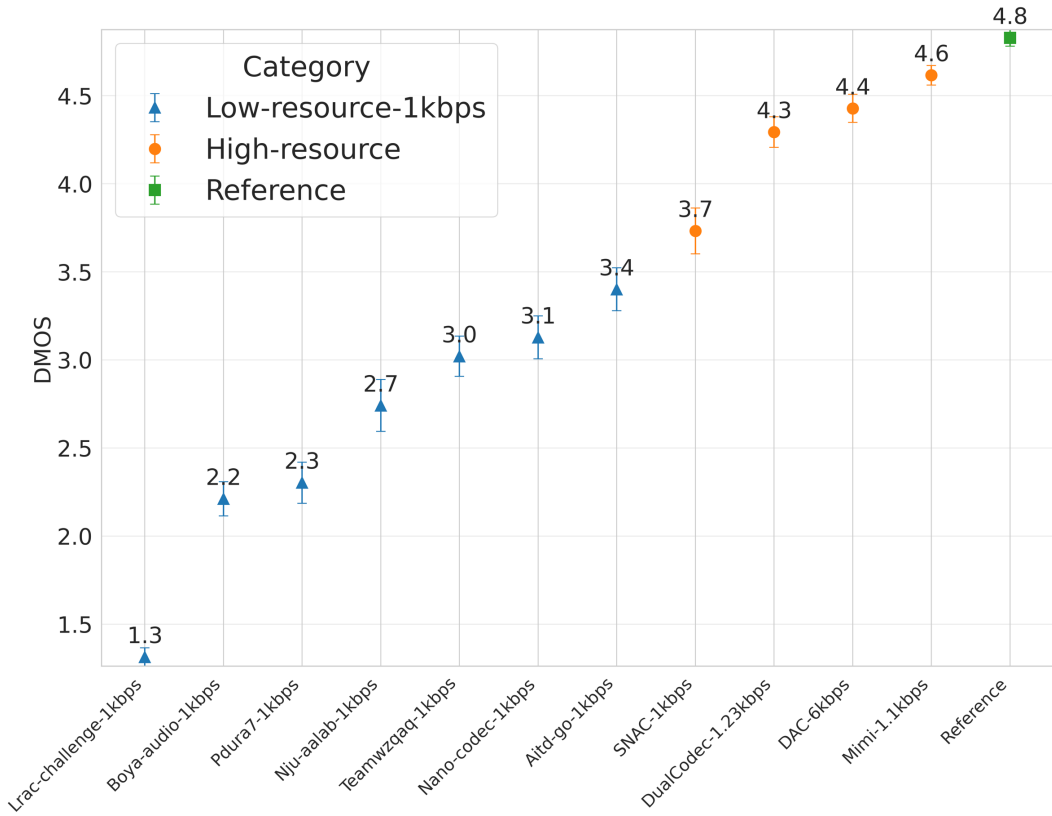


- High-resource codecs significantly outperform at ULBR
- Low-resource codecs comparable at LBR

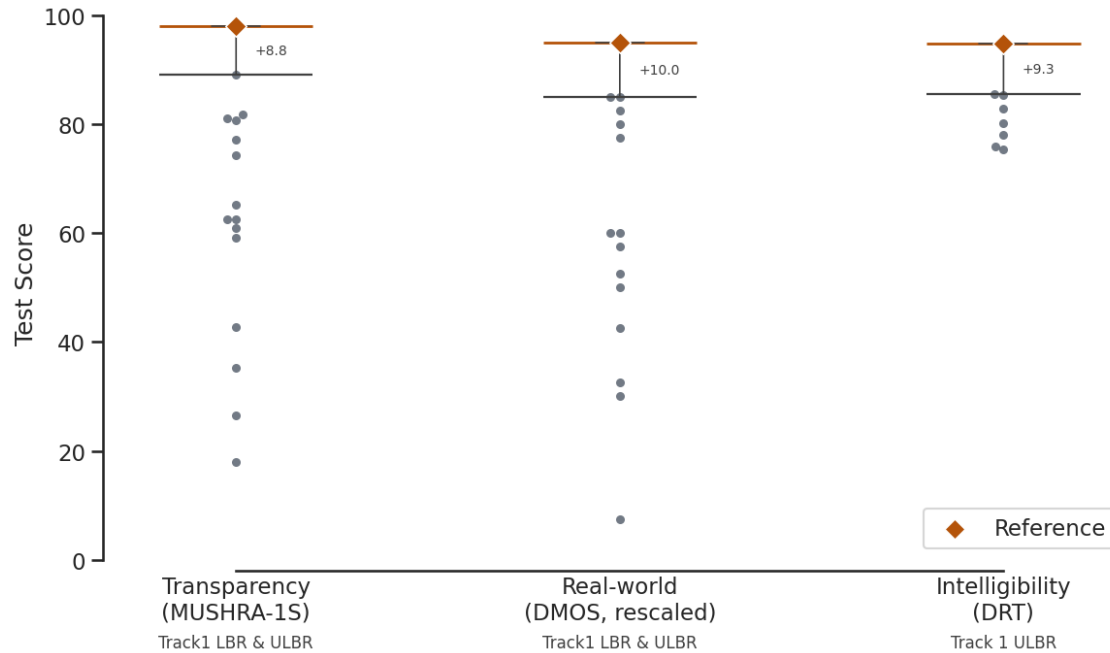


# Track 1: Real-world mild noise and reverb

- High-resource codecs significantly outperform at ULBR
- Low-resource codecs comparable at LBR



# The transparency gap



- Transparency gap for low-resource codecs
- Affects both quality and intelligibility
- Fully transparent low-resource codecs under real-world conditions are of value (even at the expense of some bitrate sacrifice)

*Intelligibility test was performed for 1 kbps challenge conditions only*

LRAC 2026  
WORKSHOP

ICASSP2026  
Barcelona

IEEE  
Signal  
Processing  
Society



# Resource constraints shaped codec design

## Common patterns across top systems

- Time-frequency representations (STFT/mel front-ends)
- Architecture:
  - Modern convolutional blocks: ConvNeXt, Conv2Former, LKCAB
  - Recurrent layers (GRU, LSTM) for temporal modeling
  - Variants of Snake activations for harmonic structure
  - Preference for joint enhancement and compression in a single network in Track 2
- Training paradigm: Generative Adversarial Networks + Multi-Stage Training:
  - **Track 1:** Multi-stage training with different weighting and combination of reconstruction, perceptual and generative losses in the stages
  - **Track 2:** Latent domain knowledge distillation + finetuning

## What's absent

- Fully transformer-based architectures and auto-regressive decoders
- Other training paradigms: Score Matching, Flow Matching, Reconstruction Only, etc.



# Quantization advancements: the road not taken

## 1. Nearly all submissions used factorized RVQ

- Alternative quantization schemes were largely absent

## 2. Differentiation came from elsewhere

- Data curation, architecture, and training drove the differences between teams

## 3. Factorized RVQ either suffices or challenging to improve

- Particularly under the constant-bitrate constraint

## 4. Open question for the field

- Is RVQ a local optimum we're stuck on, or genuinely the right tool?



# Impact of training data augmentations

## The clean-only trade-of

- nju-aalab trained on clean speech only
  - Best clean MUSHRA scores (ULBR and LBR) across all Track 1 teams
  - Dropped to 4<sup>th</sup> (ULBR)–5<sup>th</sup> (LBR) in noise/reverb robustness

## Gap due to lack of overlapping speaker augmentation

- Gap: ~1.5 DMOS points (on 5-point scale)

System	Augmentation	DMOS (simultaneous talker)
teamwzqaq, nano-codec, boya-audio	used	4.22 – 4.35
aitd-go, nju-aalab, pdura7, baseline	not used	2.05 – 2.98



# Enhancement in transparency codecs

aitd-go (PhoenixCodec) included a denoising module in Track 1 — even though it's a transparency track.

Condition	aitd-go (with enhancement)	Others (without)
ULBR noise/reverb	<b>Best</b>	Lower
LBR noise/reverb	Good	—
LBR clean	Slightly lower (reduced HF content)	Higher

- Front-end enhancement improved in-noise performance of the **PhoenixCodec** transparency codec
- Benefit was especially pronounced at 1 kbps
- At that bitrate, noise and speaker-morphing artifacts are more severe for low-resource codecs
- The benefit was at the tradeoff slight reduction of in-clean transparency performance



# Enhancement integration approaches

Four distinct paradigms emerged in Track 2:

Joint enhancement-compression codecs

## 1. End-to-end single-stage noisy-to-clean training (baseline, teamwzqaq)

- Single encoder-decoder: noisy  $\rightarrow$  clean
- Simplest but weakest

## 2. Multi-stage training with knowledge distillation in latent space (nju-aalab, xuyang, nano-codec, parslog)

- Train clean codec first  $\rightarrow$  align noisy representations to clean latent space  $\rightarrow$  retrain decoder
- **Most widely adopted by top teams**

## 3. Multi-stage training without knowledge distillation (aitd-go, boya-audio)

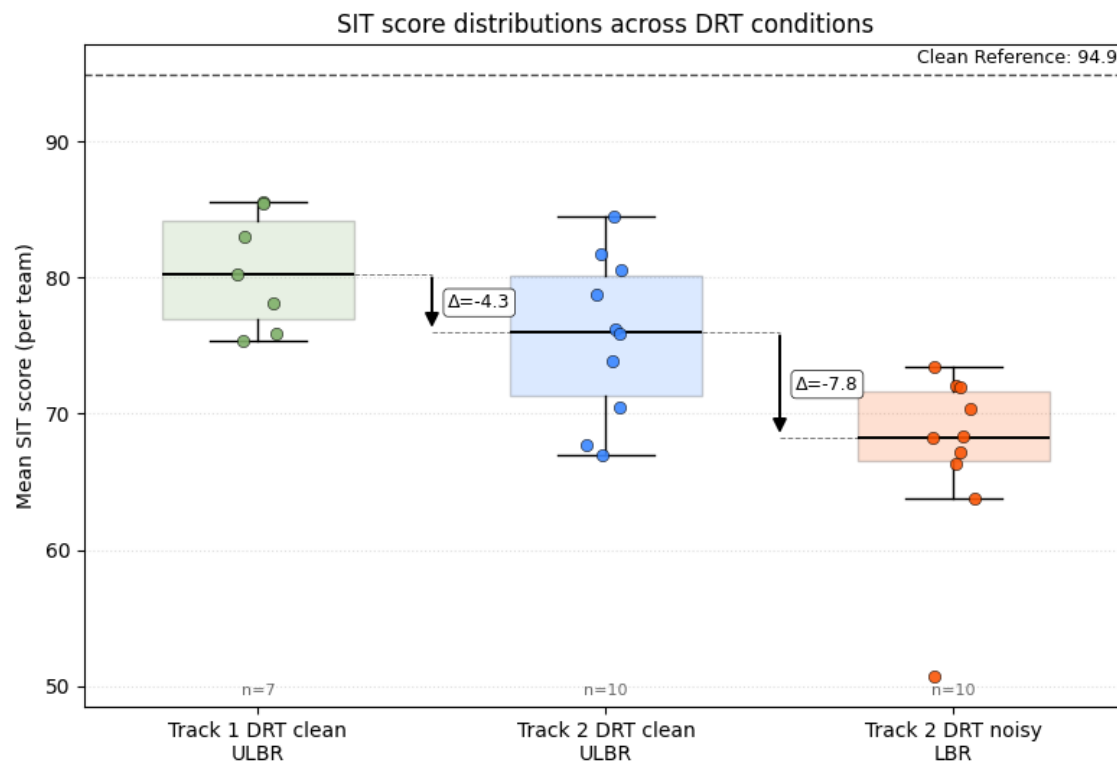
- Freeze parts of network at different stages
- Prevents catastrophic forgetting of coding quality

## 4. Cascaded architecture (leyan)

- Separate enhancement front-end  $\rightarrow$  codec
- Modular but splits latency/compute budget



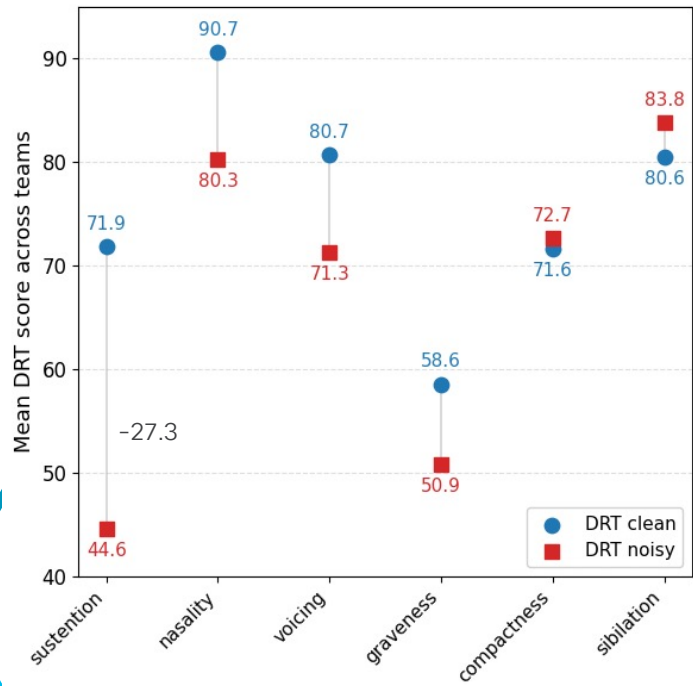
# Training for enhancement can degrade intelligibility



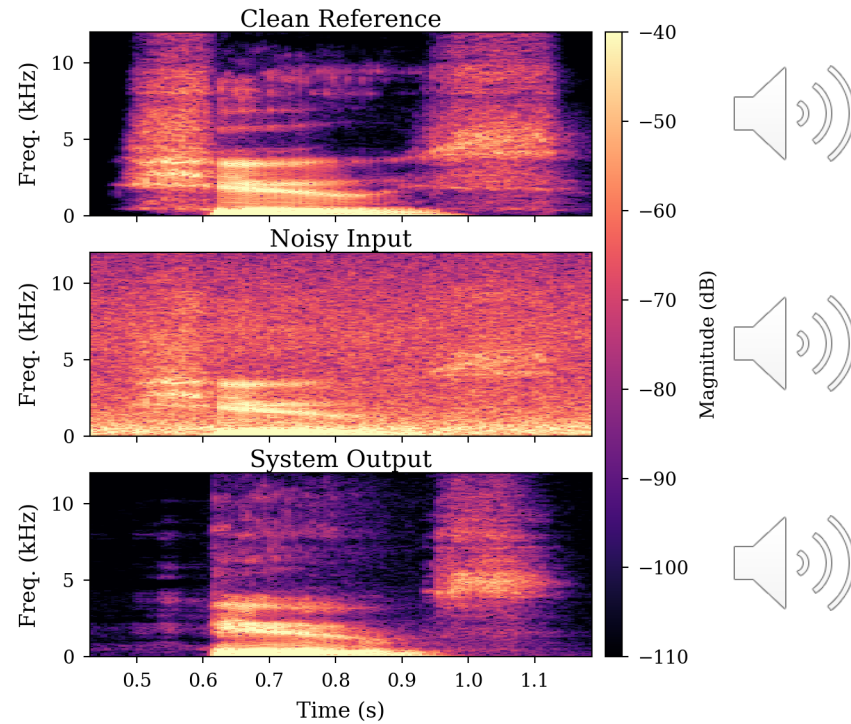
- On average, enhancement-capable codecs degraded intelligibility even in clean conditions compared to transparency codecs.



# Sustained consonants degraded the most in noisy conditions



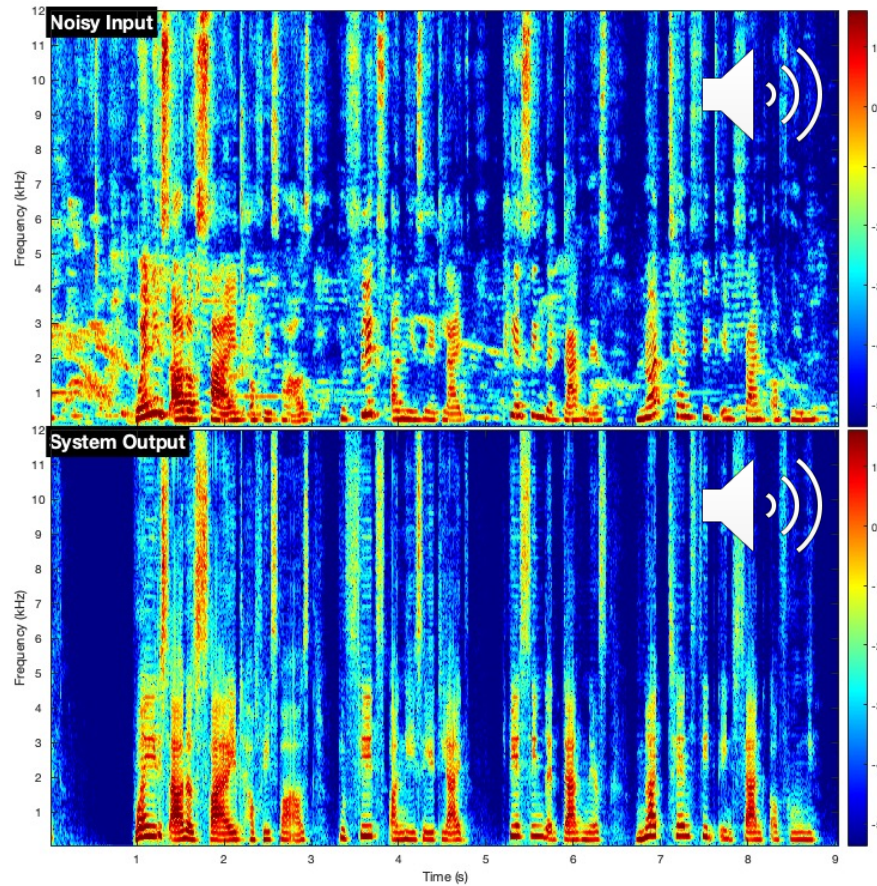
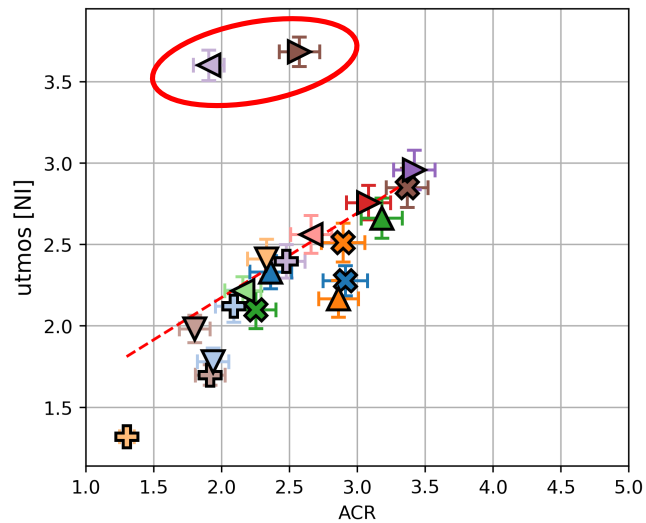
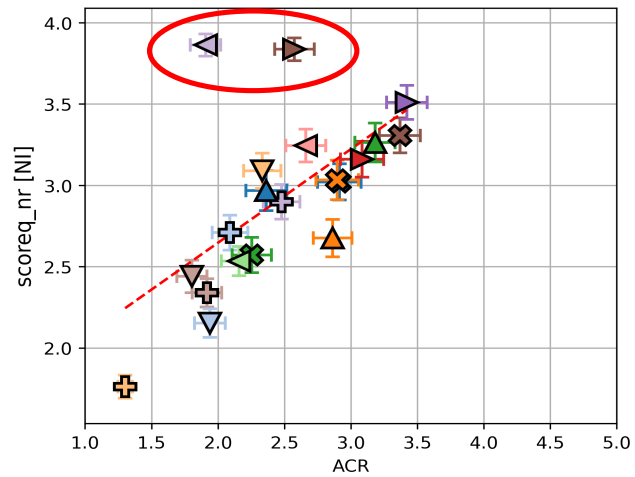
Largest clean-to-noisy degradation on sustention category



Example of sustained consonant suppression (-87 DRT score)  
/s/ "shoes" vs /tʃ/ "choose"



# Objective quality metrics may be deceived



Clean looking spectrograms but non-intelligible speech



# Evaluation insights

## What worked

- **MUSHRA-1S** — scalable single-stimulus design; context-free; effective for 17+ conditions
- **DRT** — diagnostic phonetic feature breakdown beyond aggregate intelligibility
- **Scale** — 186,720 ratings provide robust rankings

## What needs attention

- **DCR caveats** — some input files scored as low as 2.8/5; some outputs scored higher than input → listeners may have judged absolute quality rather than degradation
- Aggregate rankings remain robust, but per-file variance is high
- **Objective metrics**
  - scoreq\_ref best (Pearson 0.87) but non-intrusive metrics saturate at high quality
  - intrusive metrics (e.g., PESQ) can be unreliable for generative neural codecs

## Takeaways

- Crowdsourced listening tests at scale work.
- Objective metrics remain advisory — the gap is real and documented in our companion metrics paper.



# What's next

## Publications progress

- Outcomes paper — detailed cross-track analysis
- Objective metrics paper — extending to noisy/reverb conditions based on challenge (47 metrics × 17 codecs)

## Second edition discussions

- Keep focus on existing tracks, plus:
- Include high bitrate (16 or 24 kbps) for assessing bitrate to transparency scalability
- Mandate open-source implementation – to amplify impact, reproducibility, and help accelerate progress in this area
- We welcome your input!

## Today

- Track winners present their systems immediately after this talk
- Detailed discussion continues in the poster session and panel

**questionnaire**



# Thank You

**Thank you to all participants, reviewers, and the research community.**

Challenge website: [lrac.short.gy](http://lrac.short.gy)

Workshop website: [lrac.short.gy/workshop](http://lrac.short.gy/workshop)

Challenge description: [arXiv:2510.23312](https://arxiv.org/abs/2510.23312)

Contact: [lrac-challenge@cisco.com](mailto:lrac-challenge@cisco.com)

**Up next:** Oral presentations by the challenge track winners

- **teamwzqaq** — IRIS Codec (Track 1 Winner)
- **nju-aalab** — PR-Vocodec (Track 2 Winner)





Q&A