

## Challenge Rankings

### Rankings of Task 1:

(The rankings are determined by MOS. If the teams' MOS scores are the same, the rankings will be considered by S-MOS and N-MOS.)

Ranking	Team	MOS	S-MOS	N-MOS	dMOS	dS-MOS	dN-MOS	95%CI
1	kuaihou_deep_ns	4.02	3.87	3.87	1.46	0.94	0.84	0.02
2	HKBAT	3.86	3.78	3.80	1.30	0.85	0.77	0.02
3	WavingBrother	3.57	3.62	3.64	1.01	0.69	0.62	0.02
4	CMfeiyi	3.57	3.56	3.58	1.01	0.63	0.55	0.02
5	HNT	3.56	3.59	3.63	1.00	0.66	0.60	0.02
6	SRIB_IISC	3.45	3.38	3.36	0.70	0.32	0.25	0.04
7	GSL	3.44	3.47	3.48	0.69	0.41	0.37	0.04
.	Baseline	3.43	3.55	3.55	0.68	0.49	0.44	0.03
8	Deep Narcissus	3.34	3.47	3.49	0.59	0.41	0.39	0.04
9	Hust3iAsrLab	3.23	3.27	3.26	0.48	0.21	0.15	0.04
10	I2R-ALI	3.22	3.44	3.48	0.47	0.38	0.37	0.04
11	loiu	3.16	3.32	3.33	0.41	0.26	0.23	0.04
12	RoyalFlush	3.14	3.18	3.17	0.39	0.12	0.06	0.04
13	Doreso	2.98	3.12	3.13	0.23	0.06	0.02	0.05
14	SLADKIE	2.90	3.05	3.05	0.15	-0.01	-0.06	0.04
15	NlabAtFiveFloor	2.67	2.65	2.64	-0.08	-0.41	-0.46	0.04
	Noisy	2.56	2.93	3.03	0.00	0.00	0.00	0.02

### Rankings of Task 2:

(The rankings are determined by MOS. If the teams' MOS scores are the same, the rankings will be considered by S-MOS and N-MOS.)

Ranking	Team	MOS	S-MOS	N-MOS	dMOS	dS-MOS	dN-MOS	95%CI
1	kuaihou_deep_ns	4.14	3.93	3.92	1.63	1.05	0.93	0.02
2	HungarianDance	3.60	3.60	3.62	1.09	0.72	0.63	0.02
3	WavingBrother	3.54	3.58	3.60	1.03	0.70	0.61	0.02
.	Baseline	3.32	3.41	3.44	0.92	0.75	0.70	0.03
4	GSL	3.25	3.29	3.31	0.86	0.63	0.57	0.03
5	RoyalFlush	3.24	3.38	3.43	0.85	0.72	0.69	0.03
	Noisy	2.51	2.88	2.99	0.00	0.00	0.00	0.02

The subjective evaluation is conducted in terms of the following criteria:

**MOS:** Determination of subjective global MOS. Select the category which best describes the heard sample for purpose of everyday speech communication. The OVERALL SPEECH SAMPLE was 5-Excellent / 4-Good / 3-Fair / 2-Poor / 1-Bad.

**S-MOS:** Determination of subjective speech MOS (S-MOS). Attending ONLY to the SPEECH SIGNAL, select the category which best describes the heard sample. The SPEECH SIGNAL in this sample was 5-Not Distorted / 4-Slightly Distorted / 3-Somewhat Distorted / 2-Fairly Distorted / 1-Very Distorted.

**N-MOS:** Determination of subjective noise MOS (N-MOS). Attending ONLY to the BACKGROUND, select the category which best describes the heard sample. The BACKGROUND in this sample was 5-Not Noticeable / 4-Slightly Noticeable / 3- Noticeable But Not Intrusive / 2-Somewhat Intrusive / 1-Very Intrusive.

**dMOS/dS-MOS/dN-MOS:** Difference between the MOS / S-MOS / N-MOS after enhancement and MOS/S-MOS/N-MOS of the noisy evaluation set before enhancement.

**CI:** Confidence Interval of MOS score.

Each rater will determine MOS, S-MOS, and N-MOS scores for each evaluation audio file. And each evaluation audio file will be heard by more than 20 raters. Two rounds of subjective evaluation have been performed: 1) the first round includes submissions from all teams and 2) the second round includes several top-scoring teams for further evaluation. The final results for each task are based on the combined results

from both rounds.