

RealComm: Supplementary Result Tables

This supplement retains the detailed results behind the focused figures in the main manuscript. EERs are percentages. Values are checkpoint results from the same frozen evaluation; no additional experiments are introduced. Multiple call versions share their source utterance. Subgroup EERs do not average to the pooled EER.

TABLE I
ZERO-SHOT LANGUAGE AND DENOISING AVAILABILITY RESULTS. EN: ENGLISH; ZH: MANDARIN.

Model	Digital EN	Digital ZH	Call EN	Call ZH	Denoising unavailable
WavLM-L	5.00	3.57	39.20	40.68	33.69
Whisper	13.57	10.71	47.36	48.11	41.79
AASIST	7.14	4.29	38.10	42.26	32.86
Teffic-Audio	0.00	0.00	36.58	39.47	31.37
XLSR-SLS	35.71	41.43	51.42	49.09	48.15
DF Arena	8.57	6.43	45.24	31.58	32.80

TABLE II
DIGITAL EER BY GENERATION SYSTEM. EVERY GENERATOR USES THE SAME-DOMAIN BONA FIDE REFERENCE SET.

Model	CosyVoice2	F5-TTS	FlexiVoice	IndexTTS2	MaskGCT	Vevo2	Minimax
WavLM-L	0.00	0.00	0.00	0.00	0.00	0.00	22.50
Whisper	10.00	2.32	27.50	15.00	7.32	20.00	17.50
AASIST	2.86	0.00	14.82	5.00	2.32	9.64	7.50
Teffic-Audio	0.00	0.00	0.00	0.00	0.00	0.00	0.00
XLSR-SLS	72.50	42.50	27.50	24.82	27.50	20.18	65.36
DF Arena	6.96	0.54	0.54	2.50	0.54	0.54	32.50

TABLE III
CALL EER BY GENERATION SYSTEM. EVERY GENERATOR USES THE SAME-DOMAIN BONA FIDE REFERENCE SET.

Model	CosyVoice2	F5-TTS	FlexiVoice	IndexTTS2	MaskGCT	Vevo2	Minimax
WavLM-L	42.26	35.30	40.60	37.80	38.33	37.08	47.67
Whisper	47.50	43.81	49.88	47.37	47.32	49.94	49.70
AASIST	40.17	33.10	43.23	39.58	41.25	39.35	43.70
Teffic-Audio	40.30	33.15	37.26	36.31	34.04	35.96	47.86
XLSR-SLS	55.17	54.99	51.14	42.17	47.80	47.44	52.61
DF Arena	45.17	37.19	37.14	31.12	37.14	33.86	49.64

TABLE IV
WAVLM-L ADAPTATION EER BY CONDITION GROUP.

Subset	F	A	T	A+T
Handset	41.96	41.53	40.54	39.29
Speakerphone	45.33	42.72	42.86	41.32
LtM	23.72	21.25	18.57	17.59
Denoising off	41.47	39.92	37.44	36.51
Denoising on	39.42	37.88	35.85	34.84
Seen route	39.90	37.44	34.74	33.05
Unseen pairing	40.16	38.70	35.10	33.96
Vivo caller	42.05	40.13	37.95	36.16
Vivo receiver	38.87	38.87	34.67	34.49
English	39.20	38.96	33.72	33.49
Mandarin	40.68	38.55	36.99	34.93
Small quality loss	27.68	25.36	21.96	20.94
Medium quality loss	42.37	40.20	39.90	38.37
Large quality loss	45.46	43.83	43.37	42.30

TABLE V
AASIST ADAPTATION EER BY CONDITION GROUP.

Subset	F	A	T	A+T
Handset	45.38	42.47	43.95	41.96
Speakerphone	45.47	43.66	43.17	42.10
LtM	26.40	26.37	23.57	24.29
Denoising off	42.70	40.08	40.40	38.51
Denoising on	40.18	38.69	38.00	37.76
Seen route	38.34	36.14	36.36	35.00
Unseen pairing	39.35	38.83	38.77	38.08
Vivo caller	39.33	37.90	37.54	37.86
Vivo receiver	42.74	39.70	39.70	37.71
English	38.10	39.34	33.20	36.48
Mandarin	42.26	37.45	42.47	37.23
Small quality loss	28.72	29.16	25.89	27.30
Medium quality loss	44.18	40.64	42.73	39.92
Large quality loss	47.40	45.08	45.71	43.85

TABLE VI
TEFFIC-AUDIO ADAPTATION EER BY CONDITION GROUP.

Subset	F	A	T	A+T
Handset	41.10	40.10	29.46	27.86
Speakerphone	41.70	39.38	30.18	28.26
LtM	16.46	12.05	8.42	7.20
Denoising off	39.46	37.04	24.86	23.63
Denoising on	36.92	34.86	25.83	23.73
Seen route	34.87	32.99	22.89	20.84
Unseen pairing	38.47	35.10	23.90	21.72
Vivo caller	40.71	36.65	25.54	23.66
Vivo receiver	36.93	35.00	23.78	22.59
English	36.63	34.10	20.97	20.03
Mandarin	39.46	35.92	25.73	24.73
Small quality loss	21.53	18.39	11.51	10.97
Medium quality loss	39.13	37.78	28.11	25.28
Large quality loss	43.80	41.91	32.45	31.20