

Supplementary materials for paper "Speech Enhancement for Low Bit Rate Speech Codec"

MOS evaluation details: the total number of the ratings is $6000 = 10$ (models/techniques) \times 30 (sentences) \times 20 (times). Each listener listens to 30 sentences ($30 = 10$ models/techniques \times 3 sentences, which means each listener is able to listen to 3 audio samples for each model). The total number of the listeners is $200 = 6000/30$.

Additional results: we also provide the detailed MOS scores for VCTK and Librispeech corpus separately.

System	Codec2	VQ-VAE	Total bitrates	VCTK MOS	Librispeech MOS	Overall MOS
Original Audio (16 kHz)	-			4.07 ± 0.093	4.12 ± 0.097	4.10 ± 0.070
Opus	-		6kbps	3.34 ± 0.128	3.41 ± 0.119	3.38 ± 0.088
Codec2	-		3.2kbps	3.29 ± 0.132	3.22 ± 0.146	3.26 ± 0.098
Ours (LSGAN-V2)	1200	2400 (SC)	3.6kbps	3.57 ± 0.119	3.59 ± 0.111	3.58 ± 0.082
Ours (LSGAN-V2)	2400	2400 (SC)	4.8kbps	3.64 ± 0.111	3.70 ± 0.116	3.67 ± 0.083

TABLE I
PERFORMANCE IN TERMS OF MOS SCORE FOR THE PROPOSED AND BASELINE SYSTEMS.

ID	System	Codec2	VQ-VAE	Total bitrates	VCTK MOS	Librispeech MOS	Overall MOS
O1	Ours (w/o LSGAN)	2400	2400 (SC)	4.8kbps	3.54 ± 0.129	3.35 ± 0.132	3.44 ± 0.093
O2	Ours (LSGAN-V1)	2400	2400 (SC)	4.8kbps	3.50 ± 0.111	3.57 ± 0.121	3.53 ± 0.082
O3	Ours (LSGAN-V2)	2400	2400 (SC)	4.8kbps	3.64 ± 0.118	3.70 ± 0.116	3.67 ± 0.083

TABLE II
ABLATION STUDIES FOR EFFECTIVENESS OF THE LSGAN.

ID	System	Codec2	VQ-VAE	Total bitrates	VCTK MOS	Librispeech MOS	Overall MOS
O4	Ours (LSGAN-V2)	1200	2400 (SF)	3.6kbps	3.60 ± 0.118	3.47 ± 0.119	3.54 ± 0.082
O5	Ours (LSGAN-V2)	2400	1800 (SF)	4.2kbps	3.52 ± 0.114	3.55 ± 0.116	3.54 ± 0.084
O6	Ours (LSGAN-V2)	2400	2400 (SF)	4.8kbps	3.59 ± 0.120	3.56 ± 0.115	3.58 ± 0.083

TABLE III
PERFORMANCE IN TERMS OF MOS SCORE FOR VARIABLE BITRATES.