SLNG//

Improving Streaming STT WER Under Noise

A controlled, cross-model benchmark of streaming noise cancellation across noise types, SNR, and language.

Abstract

We benchmark four commercial streaming noise cancellation (NC) models ahead of Deepgram Nova-3 STT (ai-coustics Quail, Quail Voice Focus, NVIDIA Maxine BNR, Krisp NC), in English and Hindi, against a no-NC baseline on identical audio. Per language, 50 clips from the FLEURS corpus were mixed with six noise types, grouped into background noise and competing speech, at six SNR levels: 1,800 test conditions per language. We report micro-averaged WER with 95% bootstrap confidence intervals, pooled over the 6 to 20 dB band. Whether NC reduces WER depends on the model's optimization objective, not on the perceptual quality of its output. Only the ai-coustics models, optimized for recognition accuracy, reduce WER in their target conditions, up to 6.96 points (30% relative), with only Quail transferring to Hindi. An NC stage must therefore be validated on the target STT model, language, and noise conditions before deployment. SLNG selected the ai-coustics Quail models for the NC stage of its execution layer.

The SLNG execution layer

The SLNG execution layer is a drop-in optimization and governance layer for voice pipelines, sitting between the orchestrator and models (docs.slng.ai).

Noise cancellation runs on its STT Performance Layer.

01

Introduction

Noise cancellation (NC) removes noise from a speech-to-text (STT) model's input, to lower word error rate (WER). But perceptual quality and recognition accuracy are distinct objectives, and an NC model tuned for perceptual quality can raise WER. We measure this directly, comparing commercial NC models on a fixed streaming STT model.

The second question is whether an effect measured in English transfers to other languages. STT benchmarks are predominantly English; we test transfer in Hindi, one of the most widely spoken languages, with fast-growing adoption of voice AI.

We evaluate four commercial streaming NC models in English and Hindi, each against a no-NC baseline across six noise conditions and six signal-to-noise ratio (SNR) levels.

02

Methodology

We benchmark streaming NC ahead of a streaming STT model, measuring each model's effect on WER against a no-NC baseline on identical clean and noise-mixed audio.

2.1 STT MODEL

All audio was transcribed with Deepgram Nova-3, the same model for English and Hindi, each in its language mode. NC ran ahead of STT; the baseline sent unprocessed audio to the same model, isolating the NC's effect.

2.2 DATASET

Clips were drawn from the FLEURS multilingual read-speech corpus (Conneau et al., 2022): 50 each for English (`en_us`) and Hindi (`hi_in`), 8 to 25 words per clip, filtered to exclude proper nouns, digits, percentages, and abbreviations for normalization consistency.

2.3 NOISE CONDITIONS

Each clean clip was mixed, at each SNR level (§2.4), with two kinds of interference:

Background noise:

- Ambient: environmental background noise.
- Hum: low-frequency tone.
- Broadband noise: recorded dynamic broadband source.
- Pink noise: synthesized stationary noise (1/f spectrum).

Competing speech:

- Multi-talker: five overlapping talkers.
- Single talker: one competing talker.

The two interference types produce different errors: background noise masks target speech, causing deletions and substitutions, while competing speech is transcribed as extra words (insertions).

Competing-speech talkers were held out of the 50-clip test set in each language.

2.4 SNR CONTROL

SNR is the ratio of speech power to noise power in dB. At 0 dB speech and noise are equal (higher SNR, less noise).

Each clip-noise pair was mixed at six SNRs: 0, 3, 6, 9, 15, and 20 dB; audio was 16 kHz, mono, 16-bit. The speech level is measured over active speech: the RMS of 25 ms frames, gated with a threshold 30 dB below peak. SNR was set per clip as the ratio of this level to the noise RMS, by scaling the noise:

$$\text{gain (dB)} = \text{speech level} - \text{noise RMS} - \text{target SNR}$$

$$\text{mix} = \text{speech} + \text{noise} \cdot 10^{\text{gain} / 20}$$

Levels in dBFS, SNR in dB. Mixes peaking above -1 dBFS were scaled down, preserving the SNR.

The full grid is 50 clips × 6 noise types × 6 SNRs = 1,800 conditions per language, and the clean (no-noise) reference per clip.

We report results pooled over the **6 to 20 dB** band (determined as the production band), and exclude the 0 to 3 dB tail.

2.5 NOISE CANCELLATION MODELS

Optimized for recognition accuracy (ai-coustics states the Quail models are machine-targeted):

- ai-coustics Quail (`quail-l-16khz` , aic-sdk 0.17.0, CPU), background-noise-suppression model, at 0.25 / 0.50 / 0.75 / 1.00.
- ai-coustics Quail Voice Focus (`quail-vf-2.1-s-16khz` , aic-sdk 0.17.0, CPU), voice-isolation model, at 0.25 / 0.50 / 0.75 / 1.00 (0.80 balanced setting, ai-coustics).

General-purpose:

- NVIDIA Maxine BNR 2.0 (Background Noise Removal, Tensor Core GPU), run through the Maxine BNR NIM at 16 kHz (profile v1-16k), at 0.25 / 0.50 / 0.75 / 1.00.
- Krisp NC (Krisp VIVA UAR SDK v9.17.0, CPU), background-noise-suppression model; NSL 50 and 100 scaled to 0.50 and 1.00.

2.6 WORD ERROR RATE

We report **micro-averaged WER**, computed with jiwer (word-level Levenshtein alignment): (substitutions + deletions + insertions) divided by total reference words, pooled over the 6 to 20 dB band. Each STT transcription was scored against the clean FLEURS ground-truth transcript, both normalized first with OpenAI's EnglishTextNormalizer for English and the whisper_normalizer Hindi normalizer for Hindi. A model's effect is its change in WER versus the no-NC baseline, in percentage points (Δ pp; negative is a reduction). We break down WER into substitution, deletion, and insertion rates; the error type indicates the mechanism (§3.5).

2.7 STATISTICAL ANALYSIS

For each model and condition we bootstrap a 95% CI (1,000 resamples at the clip level, n = 50); an effect is **reliable** when its CI lies entirely above or below zero. The body reports the strongest reliable intensity; Appendix A lists all four. We report a **win / neutral / hurt** rate at each intensity (neutral: |Δ| <= 1 pp).

03

Results

NC's effect on WER divides by each model's objective: Maxine and Krisp are directionally inconsistent across noise types and SNR, while Quail Voice Focus reduces WER on English competing speech, and Quail on background noise in both languages. Effects by noise type and language follow, with full intervals in Appendix A.

3.1 BACKGROUND NOISE (ENGLISH)

Only Quail reliably reduces WER on background noise; Maxine, Krisp, and Quail Voice Focus raise it. The reduction is small and concentrated on pink and ambient noise.

MODEL	CONDITION	0.25	0.50	0.75	1.00
Quail	Background noise	-0.20*	-0.48	-0.77	**-0.94**
Quail Voice Focus	Background noise	+0.25*	+0.82*	+1.13*	+4.28
Maxine BNR	Background noise	+2.56	+2.31	+1.97	+2.40
Krisp NC	Background noise	-	+3.57	-	+3.39

*Δ WER (pp) vs no-NC, pooled over background noise; negative = reduction. Bold: strongest reliable reduction; *: CI crosses zero. Per-SNR in Appendix A.*

3.2 COMPETING SPEECH (ENGLISH)

Quail Voice Focus reduces WER the most on English competing speech.

MODEL	CONDITION	0.25	0.50	0.75	1.00
Quail	Single talker	-0.59*	-1.72	-2.79	**-2.90**
Quail	Multi-talker	-0.82*	+0.17*	+0.68*	+1.38*
Quail Voice Focus	Single talker	-2.31	-4.54	**-6.96**	-4.23
Quail Voice Focus	Multi-talker	-2.99	**-5.30**	-4.20	-2.17*
Maxine BNR	Single talker	+1.04*	-0.03*	-0.87*	-0.20*
Maxine BNR	Multi-talker	-2.93	-2.73	**-2.96**	-2.00*
Krisp NC	Single talker	-	+3.72	-	+10.03
Krisp NC	Multi-talker	-	+5.75	-	+17.98

*Δ WER (pp) vs no-NC; negative = reduction. Bold: row's strongest reliable reduction; *: CI crosses zero. Per-SNR in Appendix A.*

Quail Voice Focus's reduction is largest at 0.50 to 0.75 and smaller at full strength, where over-suppression raises the hurt rate (Appendix A). Maxine reduces only multi-talker speech; Quail reduces only single-talker speech; Krisp raises WER on both.

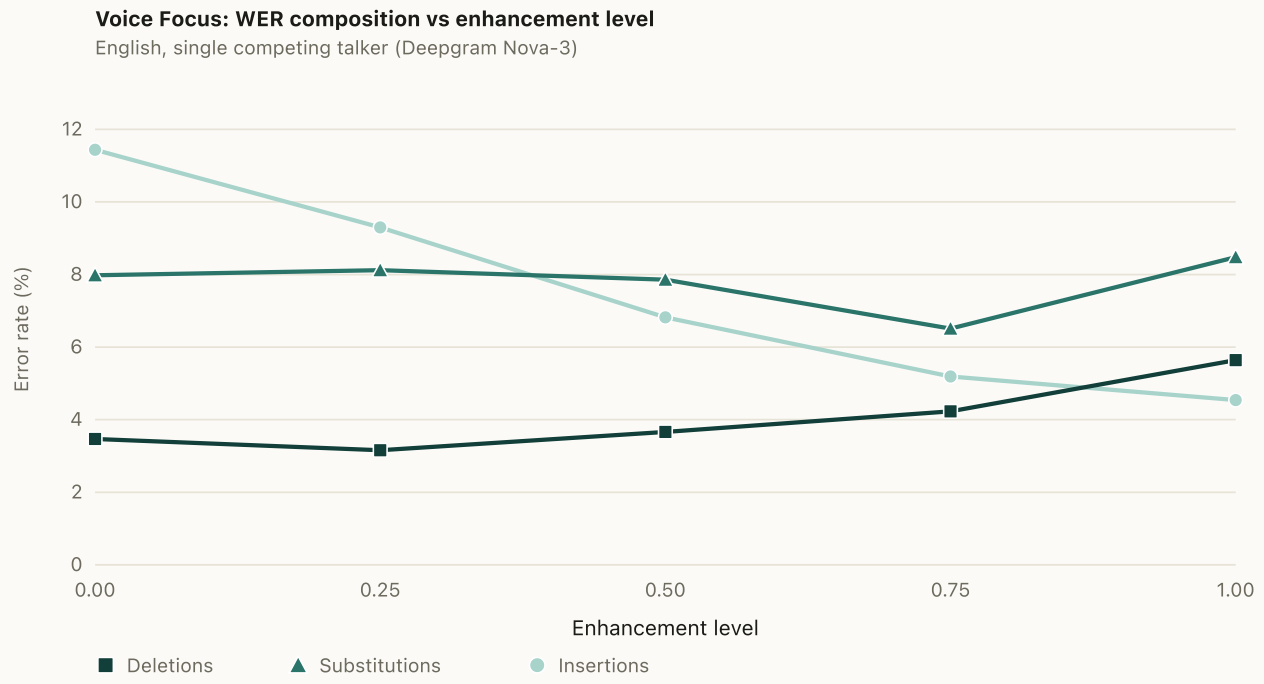

Figure 1. Quail Voice Focus on a single English competing talker: WER composition (deletions, insertions, substitutions) versus enhancement level. Insertions fall as the competing talker is removed while deletions rise as target speech is over-suppressed.

3.3 CROSS-LANGUAGE TRANSFER: HINDI

Only Quail transfers, reducing Hindi WER on background noise and marginally on competing speech. Quail Voice Focus's English competing-speech reduction flips to an increase in Hindi; Maxine and Krisp raise WER, as in English.

MODEL	CONDITION	0.25	0.50	0.75	1.00
Quail	Background noise	-0.48	-1.05	**-1.27**	-1.08
Quail	Competing speech	-0.56*	**-0.66**	-0.78*	-0.18*
Quail Voice Focus	Background noise	+1.21	+1.39	+2.50	+4.34
Quail Voice Focus	Competing speech	+0.77*	+1.87	+3.42	+6.50
Maxine BNR	Background noise	+2.91	+2.56	+3.17	+2.76
Maxine BNR	Competing speech	+0.88*	+1.59	+1.86	+1.56
Krisp NC	Background noise	-	+3.19	-	+4.29
Krisp NC	Competing speech	-	+2.26	-	+4.96

*Δ WER (pp) vs no-NC, Hindi; negative = reduction. Bold: strongest reliable reduction; *: CI crosses zero. Per-SNR in Appendix A.*

3.4 CLEAN-SPEECH PENALTY

On clean audio, only Quail stays neutral in both languages. Voice Focus raises English WER at higher intensities and Hindi WER throughout; Maxine and Krisp raise Hindi WER.

MODEL	0.25	0.50	0.75	1.00
English (no-NC 0.033)				
Quail	neutral	neutral	neutral	neutral
Voice Focus	neutral	neutral	+2.71	+3.38
Maxine BNR	neutral	neutral	neutral	neutral
Krisp NC	-	neutral	-	+1.24
Hindi (no-NC 0.136)				
Quail	neutral	neutral	neutral	neutral
Voice Focus	+1.14	+1.56	+2.81	+2.81
Maxine BNR	+2.71	+2.39	+1.98	+1.46
Krisp NC	-	+2.81	-	+2.29

Clean-speech penalty, mean Δ pp vs no-NC. neutral = |Δ| <= 1 pp.

3.5 ERROR STRUCTURE

On background noise, WER changes are deletion-driven (Figure 2). Voice Focus's English competing-speech reduction is insertion-driven: the STT transcribes the interferer as extra words (Figure 1). Hindi produces fewer insertions; Voice Focus's reduction does not transfer. Quail's deletion-driven reduction transfers to Hindi.

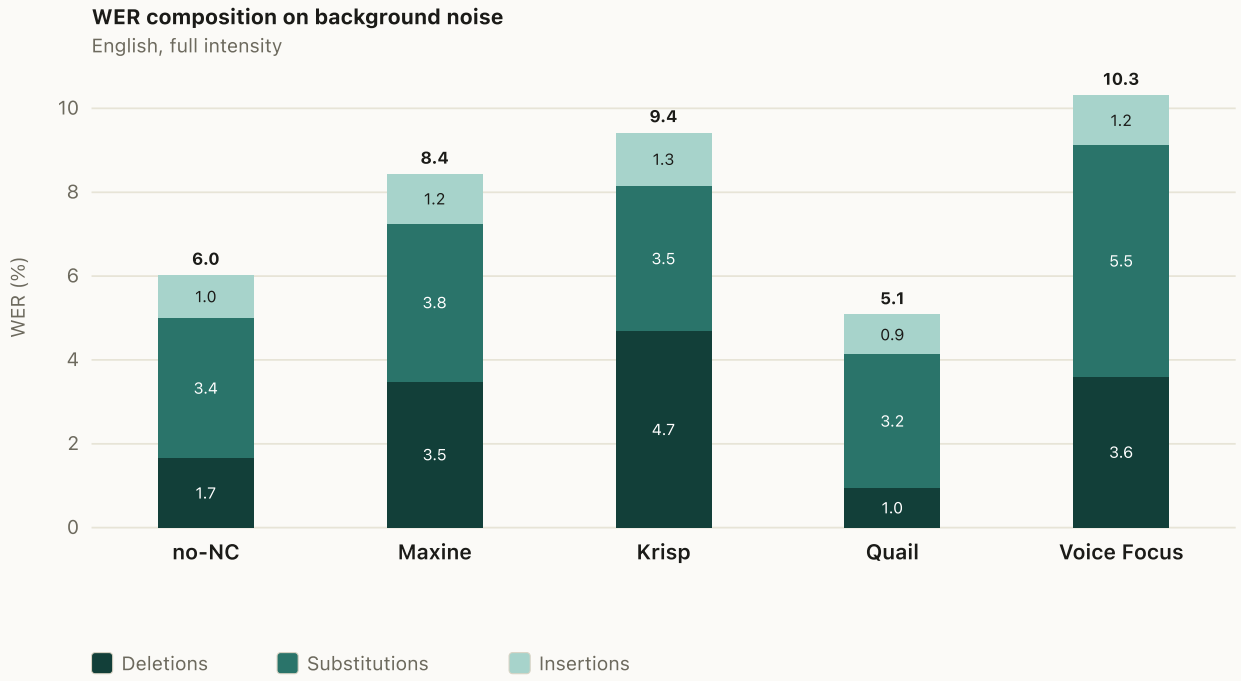

Figure 2. WER composition on background noise, by error type (English, full intensity).

04

Conclusion

Whether NC reduces WER depends on the model's optimization objective, not on the perceptual quality of its output. No model reduces WER across all conditions (Appendix A). The measured effects are specific to the model, noise type, language, and intensity setting, and an effect measured in one language cannot be assumed in another (§3.3). An unvalidated NC stage risks the WER increases documented in §3, consistent with LiveKit's report that some STT models perform better without NC (LiveKit). Deploying NC ahead of an STT model therefore requires a WER sweep over candidate models and intensities on the target STT model, language, and noise conditions.

An NC stage ahead of a streaming STT model must be:

- Optimized against WER. It must suppress interference while retaining target speech (§3.5).
- Reliable in target conditions. It must reduce WER for background noise or competing speech across SNR levels, and remain neutral otherwise, with a low per-clip hurt rate (§2.7).
- Validated per language. Cross-language transfer cannot be assumed (§3.3); WER must be measured per deployment language.
- Streaming-native. It must run causally, within the latency headroom of real-time streaming.
- Compute-efficient. Its per-stream cost must not bottleneck STT concurrency.

No model we tested meets all of these requirements. Streaming NC that reduces WER across noise conditions and languages remains an open problem.

On these results, SLNG is integrating the ai-coustics Quail models as the NC stage of its execution layer, validated per language and condition.

Contributors

Contributors are listed in alphabetical order by last name.

Jiameng Gao, Alfredo Ibañez, Ismael Ordaz Reyes, Nick Simmonds, Dmitry Skorinko, Tolga Yapici

Appendix

A. FULL RESULTS

A.1 ENGLISH

Quail

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.033	0.038 +0.56 [-0.23, +1.29] 8/74/18	0.037 +0.45 [-0.23, +1.10] 6/80/14	0.038 +0.56 [-0.24, +1.32] 4/80/16	0.029 -0.34 [-1.55, +0.77] 14/72/14
Ambient	0.069	0.063 -0.62 [-1.43, +0.19] 20/68/12	0.063 -0.59 [-1.53, +0.36] 27/58/15	0.060 -0.93 [-1.77, -0.05] 26/60/13	0.057 -1.21 [-2.04, -0.37] 28/58/14
Hum	0.035	0.032 -0.25 [-0.97, +0.39] 16/70/14	0.033 -0.17 [-0.85, +0.52] 16/67/16	0.033 -0.14 [-0.94, +0.59] 18/63/19	0.033 -0.17 [-0.91, +0.57] 18/62/20
Broadband noise	0.044	0.046 +0.23 [-0.41, +1.02] 10/78/12	0.048 +0.37 [-0.41, +1.20] 12/74/14	0.044 -0.03 [-0.72, +0.71] 14/73/13	0.041 -0.34 [-1.07, +0.39] 18/68/14
Pink noise	0.094	0.092 -0.17 [-0.96, +0.78] 20/60/20	0.078 -1.52 [-2.66, -0.25] 27/52/20	0.074 -1.97 [-3.09, -0.83] 30/51/20	0.073 -2.06 [-3.24, -0.82] 34/46/20
Multi-talker	0.188	0.180 -0.82 [-2.44, +0.91] 29/49/22	0.190 +0.17 [-1.80, +2.18] 30/44/26	0.195 +0.68 [-1.53, +3.04] 32/36/32	0.202 +1.38 [-0.90, +3.92] 33/34/34
Single talker	0.229	0.223 -0.59 [-1.61, +0.66] 26/54/21	0.212 -1.72 [-2.97, -0.40] 32/47/22	0.201 -2.79 [-4.06, -1.54] 38/44/18	0.200 -2.90 [-4.55, -1.35] 39/42/19

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.033	+0.56	+0.45	+0.56	-0.34
Ambient	0	0.487	-4.06	-8.91	-11.05	-10.94
	3	0.211	-0.34	+0.00	+0.45	+0.00
	6	0.120	-0.56	-1.47	-2.48	-2.25
	9	0.072	-1.35	-1.24	-1.80	-1.92
	15	0.050	-1.13	-1.58	-1.24	-1.69
	20	0.035	+0.56	+1.92	+1.80	+1.01
Hum	0	0.057	-0.45	-0.90	-1.24	-1.58
	3	0.048	-0.90	-1.80	-1.13	-1.13
	6	0.045	-1.13	-1.47	-1.24	-1.13
	9	0.034	-0.23	+0.11	-0.34	-0.23
	15	0.032	+0.11	+0.11	+0.00	+0.11
	20	0.028	+0.23	+0.56	+1.01	+0.56

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Broadband noise	0	0.123	-0.34	-1.47	-0.34	-0.23
	3	0.070	+1.69	+2.03	+1.13	+1.47
	6	0.063	+0.11	+0.68	+0.45	-0.11
	9	0.047	+0.68	+0.90	-0.11	-0.68
	15	0.037	+0.11	-0.45	-0.68	-0.68
	20	0.028	+0.00	+0.34	+0.23	+0.11
Pink noise	0	0.331	-2.93	-5.98	-6.31	-6.88
	3	0.218	-0.79	-1.92	-3.95	-4.06
	6	0.157	-1.24	-3.49	-3.72	-3.61
	9	0.100	+1.58	-0.90	-1.92	-2.14
	15	0.067	-0.11	-0.90	-1.92	-1.47
	20	0.051	-0.90	-0.79	-0.34	-1.01
Multi-talker	0	0.820	+2.14	+1.92	+2.82	+3.83
	3	0.619	+1.80	+2.03	+2.37	+1.58
	6	0.392	-1.24	+1.58	+3.38	+4.40
	9	0.221	-2.48	-1.47	-0.56	+0.34
	15	0.079	-0.11	+0.68	-0.34	+0.90
	20	0.060	+0.56	-0.11	+0.23	-0.11
Single talker	0	0.690	-0.79	-2.71	-0.79	-3.27
	3	0.526	-3.49	-4.62	-4.51	-6.76
	6	0.365	+0.23	-1.80	-1.47	-1.69
	9	0.284	-1.58	-2.59	-4.74	-4.85
	15	0.147	+0.45	-0.11	-2.14	-2.14
	20	0.120	-1.47	-2.37	-2.82	-2.93

Quail Voice Focus

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.033	0.034 +0.11 [-1.00, +1.24] 14/70/16	0.039 +0.68 [-0.93, +2.47] 14/70/16	0.060 +2.71 [-0.83, +7.51] 14/76/10	0.067 +3.38 [-0.33, +8.90] 14/64/22
Ambient	0.069	0.079 +1.07 [-0.33, +2.64] 21/59/20	0.096 +2.68 [+0.17, +6.16] 20/56/23	0.086 +1.75 [+0.09, +3.65] 24/51/26	0.125 +5.61 [+2.59, +9.24] 20/44/37
Hum	0.035	0.045 +0.99 [-0.17, +2.58] 14/68/18	0.036 +0.17 [-0.54, +1.03] 16/68/15	0.051 +1.63 [-0.06, +4.08] 17/66/17	0.070 +3.52 [+0.17, +8.26] 20/54/26
Broadband noise	0.044	0.049 +0.54 [-0.39, +1.59] 12/72/16	0.053 +0.90 [-0.20, +2.02] 14/67/18	0.060 +1.63 [+0.14, +3.41] 14/65/22	0.107 +6.26 [+2.85, +10.37] 14/48/38
Pink noise	0.094	0.078 -1.58 [-3.13, -0.09] 28/56/16	0.089 -0.45 [-2.16, +1.37] 28/49/22	0.089 -0.51 [-2.15, +1.13] 30/45/26	0.111 +1.75 [-1.18, +4.79] 28/41/32
Multi-talker	0.188	0.158 -2.99 [-5.72, -0.35] 40/42/18	0.135 -5.30 [-7.98, -2.87] 40/41/18	0.146 -4.20 [-7.74, -0.38] 46/27/27	0.167 -2.17 [-6.06, +1.42] 44/26/30

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Single talker	0.229	0.206 -2.31 [-4.28, -0.44] 40/38/21	0.183 -4.54 [-6.78, -2.32] 50/32/18	0.159 -6.96 [-9.30, -4.66] 62/24/14	0.187 -4.23 [-7.37, -0.70] 54/24/22

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.033	+0.11	+0.68	+2.71	+3.38
Ambient	0	0.490	-3.95	-1.24	+0.56	+12.63
	3	0.211	+2.37	+3.38	+6.65	+16.35
	6	0.118	+1.80	+2.59	+3.83	+11.39
	9	0.072	+1.47	+1.13	+1.24	+4.85
	15	0.050	-0.79	+2.03	-0.23	+1.80
	20	0.035	+1.80	+4.96	+2.14	+4.40
Hum	0	0.057	-1.13	+0.11	+2.14	+4.85
	3	0.050	-0.11	+0.56	+0.56	+3.16
	6	0.045	+0.11	-0.68	+0.90	+3.38
	9	0.034	+0.90	+0.79	+1.58	+3.49
	15	0.032	+2.37	+0.11	+1.58	+3.38
	20	0.028	+0.56	+0.45	+2.48	+3.83
Broadband noise	0	0.123	-0.11	+0.34	+3.83	+18.71
	3	0.070	+2.48	+3.83	+4.28	+14.09
	6	0.063	+0.56	+0.90	+0.90	+9.36
	9	0.047	+0.56	+1.58	+2.25	+7.67
	15	0.037	+0.23	+0.23	+1.47	+3.49
	20	0.028	+0.79	+0.90	+1.92	+4.51
Pink noise	0	0.331	-7.10	-5.64	+1.69	+9.13
	3	0.218	-0.11	-1.69	+2.25	+5.98
	6	0.157	-3.95	-3.72	-0.79	+1.69
	9	0.100	-2.25	-0.23	+0.11	+3.95
	15	0.067	+0.34	+2.03	-0.68	+0.11
	20	0.051	-0.45	+0.11	-0.68	+1.24
Multi-talker	0	0.820	-1.35	+1.24	+2.59	+6.88
	3	0.619	-2.25	-3.38	-3.49	+13.42
	6	0.393	-3.95	-10.60	-10.71	-5.98
	9	0.221	-7.22	-8.91	-8.34	-5.19
	15	0.079	-2.14	-0.45	+0.68	+1.80
	20	0.060	+1.35	-1.24	+1.58	+0.68
Single talker	0	0.690	+0.00	-1.80	-4.06	-3.16
	3	0.526	-0.56	-4.74	-4.40	-4.62
	6	0.365	-0.23	-2.48	-7.22	-2.71
	9	0.284	-4.62	-5.98	-9.13	-7.33
	15	0.147	-1.35	-3.95	-5.41	-3.49
	20	0.120	-3.04	-5.75	-6.09	-3.38

Maxine BNR

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.033	0.034 +0.11 [-0.93, +1.13] 12/72/16	0.038 +0.56 [-0.76, +1.86] 16/60/24	0.035 +0.23 [-0.91, +1.37] 16/64/20	0.036 +0.34 [-0.83, +1.56] 14/68/18
Ambient	0.069	0.096 +2.71 [+0.96, +4.87] 14/56/29	0.089 +1.97 [+0.59, +3.55] 15/60/26	0.092 +2.31 [+0.70, +4.22] 17/57/26	0.098 +2.93 [+1.14, +4.97] 14/59/27
Hum	0.035	0.056 +2.17 [+0.73, +4.05] 15/56/28	0.054 +1.97 [+0.54, +3.93] 16/57/26	0.048 +1.30 [+0.33, +2.51] 18/57/25	0.054 +1.89 [+0.76, +3.26] 14/56/30
Broadband noise	0.044	0.081 +3.72 [+1.84, +5.91] 12/57/30	0.076 +3.24 [+1.58, +5.10] 12/56/32	0.070 +2.56 [+1.20, +4.08] 12/58/30	0.067 +2.34 [+1.10, +3.84] 12/56/31
Pink noise	0.094	0.110 +1.66 [+0.42, +2.97] 16/52/32	0.114 +2.06 [+0.81, +3.45] 18/50/32	0.111 +1.72 [+0.13, +3.72] 18/52/30	0.118 +2.45 [+0.80, +4.31] 18/48/34
Multi-talker	0.188	0.159 -2.93 [-4.91, -1.12] 41/36/24	0.161 -2.73 [-5.08, -0.48] 38/36/25	0.159 -2.96 [-5.18, -0.88] 42/36/22	0.168 -2.00 [-4.31, +0.22] 42/32/26
Single talker	0.229	0.239 +1.04 [-0.40, +2.66] 28/35/37	0.229 -0.03 [-1.42, +1.57] 30/36/33	0.220 -0.87 [-2.73, +0.99] 34/34/32	0.227 -0.20 [-2.10, +1.80] 32/34/33

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.033	+0.11	+0.56	+0.23	+0.34
Ambient	0	0.490	+2.25	+1.92	-0.11	+2.93
	3	0.211	+4.74	+5.64	+5.52	+5.86
	6	0.118	+3.83	+3.27	+2.03	+2.71
	9	0.072	+1.13	+0.23	+0.34	+1.35
	15	0.050	+4.85	+4.17	+5.30	+4.85
	20	0.035	+1.01	+0.23	+1.58	+2.82
Hum	0	0.057	-0.34	-0.23	-0.56	+0.56
	3	0.050	-0.56	+0.34	-0.11	-0.23
	6	0.045	+0.34	+0.45	-0.23	+1.01
	9	0.034	+1.24	+0.45	+0.68	+0.90
	15	0.032	+3.72	+4.74	+0.68	+3.72
	20	0.028	+3.38	+2.25	+4.06	+1.92
Broadband noise	0	0.123	+2.25	+2.25	+2.14	+2.82
	3	0.070	+4.17	+3.72	+3.27	+3.49
	6	0.063	+2.93	+1.92	+2.14	+1.92
	9	0.047	+1.47	+1.80	+1.92	+1.35
	15	0.037	+6.20	+6.20	+2.48	+3.04
	20	0.028	+4.28	+3.04	+3.72	+3.04
Pink noise	0	0.331	+1.69	+0.34	+2.59	+0.79
	3	0.218	-0.34	+0.23	-0.11	+1.35
	6	0.157	+2.37	+2.25	+1.58	+0.79
	9	0.100	+0.90	-0.11	+0.11	+1.47

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
	15	0.067	+2.03	+4.17	+1.80	+1.69
	20	0.051	+1.35	+1.92	+3.38	+5.86
Multi-talker	0	0.820	-3.27	-2.59	-4.17	-2.03
	3	0.619	-6.31	-6.54	-5.52	-5.52
	6	0.393	-8.12	-9.81	-9.47	-7.10
	9	0.221	-4.17	-3.16	-4.51	-3.95
	15	0.079	-0.34	-0.11	+2.25	-0.23
	20	0.060	+0.90	+2.14	-0.11	+3.27
Single talker	0	0.690	+5.30	+2.59	+2.48	+2.71
	3	0.526	+1.80	+0.23	+1.58	+0.00
	6	0.365	+4.74	+1.80	+1.47	+2.25
	9	0.284	-1.69	-0.34	-3.27	-2.59
	15	0.147	+2.03	+0.23	+0.45	+0.34
	20	0.120	-0.90	-1.80	-2.14	-0.79

Krisp NC

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	NSL 50	NSL 100
Clean	0.033	0.025 -0.79 [-1.94, +0.23] 16/74/10	0.045 +1.24 [-1.56, +5.91] 16/72/12
Ambient	0.069	0.103 +3.38 [+0.36, +7.46] 22/57/20	0.120 +5.16 [+1.81, +9.27] 24/44/32
Hum	0.035	0.082 +4.74 [+1.37, +8.52] 16/66/18	0.060 +2.54 [-0.38, +6.35] 20/62/18
Broadband noise	0.044	0.097 +5.30 [+1.42, +9.53] 14/67/18	0.087 +4.28 [+1.01, +8.16] 16/57/26
Pink noise	0.094	0.102 +0.85 [-2.14, +4.05] 32/49/18	0.109 +1.58 [-2.14, +5.48] 40/44/16
Multi-talker	0.188	0.246 +5.75 [+1.13, +10.65] 34/31/36	0.368 +17.98 [+12.37, +24.37] 22/18/60
Single talker	0.229	0.266 +3.72 [+0.36, +7.01] 28/38/34	0.329 +10.03 [+4.09, +16.70] 34/18/48

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	NSL 50	NSL 100
Clean		0.033	-0.79	+1.24
Ambient	0	0.490	-8.23	+1.35
	3	0.211	+0.90	+10.26

NOISE	SNR	BASELINE WER	NSL 50	NSL 100
	6	0.118	+0.23	+4.74
	9	0.072	+0.00	+2.82
	15	0.050	+6.43	+5.75
	20	0.035	+6.88	+7.33
Hum	0	0.057	+1.92	+0.11
	3	0.050	-0.45	+1.69
	6	0.045	-1.24	-0.68
	9	0.034	+0.00	-0.34
	15	0.032	+10.82	+5.41
	20	0.028	+9.36	+5.75
Broadband noise	0	0.123	-0.68	+6.31
	3	0.070	+1.69	+5.64
	6	0.063	-0.23	+2.59
	9	0.047	+1.13	+1.69
	15	0.037	+9.47	+7.78
	20	0.028	+10.82	+5.07
Pink noise	0	0.331	-8.34	-0.68
	3	0.218	-5.41	-4.17
	6	0.157	-4.85	-4.62
	9	0.100	+1.92	-1.47
	15	0.067	+1.01	+3.49
	20	0.051	+5.30	+8.91
Multi-talker	0	0.820	-0.90	+13.64
	3	0.619	-0.11	+13.19
	6	0.393	-0.79	+14.66
	9	0.221	-1.13	+12.74
	15	0.079	+10.15	+24.35
	20	0.060	+14.77	+20.18
Single talker	0	0.690	+1.35	+3.27
	3	0.526	+1.80	+6.31
	6	0.365	+5.41	+10.26
	9	0.284	+4.17	+8.12
	15	0.147	+3.72	+12.74
	20	0.120	+1.58	+9.02

A.2 HINDI

Quail

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.136	0.144 +0.73 [-0.33, +1.72] 6/74/20	0.142 +0.52 [-0.59, +1.61] 10/70/20	0.138 +0.21 [-0.92, +1.32] 12/68/20	0.138 +0.21 [-0.82, +1.20] 12/70/18
Ambient	0.192	0.185 -0.65 [-1.62, +0.31] 22/60/18	0.181 -1.09 [-2.20, -0.05] 25/55/20	0.182 -1.01 [-2.08, +0.00] 24/57/19	0.184 -0.75 [-1.70, +0.22] 26/52/21

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Hum	0.160	0.149 -1.12 [-1.89, -0.40] 30/56/14	0.149 -1.17 [-1.88, -0.54] 28/60/12	0.145 -1.51 [-2.43, -0.69] 33/54/14	0.148 -1.20 [-2.18, -0.24] 34/49/17
Broadband noise	0.165	0.165 -0.05 [-1.01, +0.84] 19/57/24	0.164 -0.10 [-1.08, +0.82] 20/55/24	0.166 +0.10 [-1.12, +1.36] 24/48/28	0.174 +0.88 [-0.31, +2.13] 20/47/33
Pink noise	0.210	0.209 -0.10 [-1.10, +0.83] 17/62/22	0.192 -1.82 [-3.88, -0.20] 26/51/22	0.183 -2.65 [-4.51, -0.95] 32/48/20	0.177 -3.25 [-5.48, -1.45] 36/43/21
Multi-talker	0.169	0.166 -0.29 [-1.05, +0.41] 20/64/17	0.165 -0.42 [-1.22, +0.28] 26/54/20	0.163 -0.55 [-1.49, +0.42] 30/49/22	0.163 -0.55 [-1.50, +0.49] 31/48/21
Single talker	0.217	0.209 -0.83 [-1.80, +0.00] 23/57/20	0.208 -0.91 [-1.99, +0.05] 26/52/23	0.207 -1.01 [-2.13, +0.11] 28/50/22	0.219 +0.18 [-1.08, +1.51] 26/46/28

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.136	+0.73	+0.52	+0.21	+0.21
Ambient	0	0.556	-5.52	-9.57	-13.01	-16.23
	3	0.343	-3.12	-3.43	-5.52	-6.04
	6	0.270	-2.81	-3.75	-3.85	-3.23
	9	0.201	+0.00	-1.04	-1.04	-1.35
	15	0.156	+0.10	+0.21	+0.00	+0.21
	20	0.140	+0.10	+0.21	+0.83	+1.35
Hum	0	0.201	-1.56	-2.39	-3.43	-3.95
	3	0.178	-0.42	-1.46	-1.25	-2.08
	6	0.162	-0.52	-0.94	-1.04	-0.83
	9	0.159	-0.83	-0.94	-1.35	-0.10
	15	0.163	-1.98	-1.66	-2.19	-2.39
	20	0.156	-1.14	-1.14	-1.46	-1.46
Broadband noise	0	0.286	-1.25	-4.68	-4.47	-5.83
	3	0.260	-5.10	-5.72	-6.14	-5.83
	6	0.200	-1.66	-1.35	-1.04	+0.62
	9	0.171	+0.94	+0.31	+0.42	+1.14
	15	0.150	+0.00	+0.42	+0.52	+0.83
	20	0.140	+0.52	+0.21	+0.52	+0.94
Pink noise	0	0.504	-6.04	-9.37	-11.45	-12.59
	3	0.370	-2.50	-4.58	-4.89	-6.14
	6	0.296	-0.62	-4.79	-5.31	-6.66
	9	0.232	-0.83	-3.12	-4.68	-5.20
	15	0.163	+0.83	+0.52	-0.31	-1.14
	20	0.149	+0.21	+0.10	-0.31	+0.00
Multi-talker	0	0.402	-3.23	-3.23	-6.35	-4.58
	3	0.289	-2.19	-1.66	-3.23	-3.23
	6	0.204	-0.31	-1.25	-0.94	-0.21
	9	0.169	+0.10	+0.42	-0.21	-0.10

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
	15	0.157	-0.94	-0.83	-1.04	-1.77
	20	0.146	+0.00	+0.00	+0.00	-0.10
Single talker	0	0.543	-0.52	+0.31	+0.10	-0.10
	3	0.385	-1.35	-1.35	-2.19	-2.29
	6	0.305	-3.95	-4.27	-4.47	-2.39
	9	0.207	+0.52	+0.42	+0.31	+2.08
	15	0.173	+1.04	+0.73	+0.52	+1.25
	20	0.185	-0.94	-0.52	-0.42	-0.21

Quail Voice Focus

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.136	0.148 +1.14 [-0.31, +3.00] 12/66/22	0.152 +1.56 [+0.10, +3.40] 10/62/28	0.164 +2.81 [+0.52, +5.81] 16/50/34	0.164 +2.81 [+0.54, +6.18] 18/48/34
Ambient	0.192	0.205 +1.30 [-0.03, +2.90] 22/52/27	0.210 +1.77 [+0.25, +3.35] 21/48/30	0.221 +2.91 [+1.00, +5.08] 20/44/36	0.243 +5.10 [+2.76, +7.97] 18/38/45
Hum	0.160	0.163 +0.26 [-1.33, +2.17] 30/48/22	0.172 +1.14 [-0.76, +3.62] 26/48/25	0.180 +1.98 [-0.78, +5.33] 31/42/28	0.189 +2.86 [+0.24, +6.37] 28/40/33
Broadband noise	0.165	0.185 +1.98 [+0.38, +3.99] 20/51/28	0.189 +2.42 [+0.55, +4.85] 18/46/36	0.204 +3.95 [+1.09, +7.62] 21/44/35	0.225 +5.96 [+2.57, +10.07] 19/36/46
Pink noise	0.211	0.224 +1.30 [-0.21, +3.09] 24/48/28	0.213 +0.21 [-1.17, +1.87] 26/47/26	0.222 +1.14 [-0.98, +3.46] 28/38/34	0.245 +3.43 [+0.82, +6.24] 30/31/39
Multi-talker	0.169	0.181 +1.17 [-0.50, +3.30] 24/52/24	0.187 +1.80 [-0.13, +4.36] 24/47/28	0.209 +4.06 [+1.16, +8.09] 22/41/36	0.241 +7.26 [+3.70, +11.88] 18/36/45
Single talker	0.220	0.223 +0.36 [-1.15, +2.06] 27/46/27	0.239 +1.95 [+0.23, +3.92] 24/48/28	0.248 +2.78 [+0.86, +5.05] 22/44/34	0.277 +5.75 [+3.00, +9.35] 20/32/49

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.136	+1.14	+1.56	+2.81	+2.81
Ambient	0	0.554	-0.31	-1.56	+3.02	+17.17
	3	0.343	+1.35	+4.68	+2.19	+8.12
	6	0.270	-0.94	-0.73	+0.73	+6.24
	9	0.201	+1.66	+1.66	+2.39	+5.31
	15	0.157	+2.60	+1.35	+4.89	+3.95
	20	0.140	+1.87	+4.79	+3.64	+4.89
Hum	0	0.202	+0.52	+1.25	+7.18	+12.28
	3	0.179	+1.66	+1.04	+2.50	+7.39

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
	6	0.162	+0.73	+1.46	+0.94	+3.43
	9	0.159	+0.31	+1.35	+2.50	+3.85
	15	0.163	-0.62	+0.21	+3.23	+1.46
	20	0.156	+0.62	+1.56	+1.25	+2.71
Broadband noise	0	0.287	-1.04	-1.77	+0.73	+8.01
	3	0.260	-3.43	-2.08	-0.62	+7.80
	6	0.200	+0.21	+1.25	+2.39	+6.56
	9	0.171	+2.08	+2.60	+4.27	+4.89
	15	0.149	+2.71	+2.50	+4.27	+6.35
	20	0.140	+2.91	+3.33	+4.89	+6.04
Pink noise	0	0.504	-6.24	-10.41	-7.39	+1.25
	3	0.370	-1.46	-4.16	-2.81	+1.56
	6	0.294	-0.73	-1.66	-1.25	+2.71
	9	0.233	+1.46	-1.35	-0.10	+3.95
	15	0.165	+1.77	+1.66	+1.35	+3.23
	20	0.151	+2.71	+2.19	+4.58	+3.85
Multi-talker	0	0.402	+1.25	+2.08	+5.31	+16.86
	3	0.289	+2.60	+4.37	+2.71	+9.78
	6	0.204	+1.35	+1.04	+5.62	+10.09
	9	0.169	+1.25	+2.08	+4.27	+7.28
	15	0.157	-0.73	+0.73	+2.08	+5.41
	20	0.146	+2.81	+3.33	+4.27	+6.24
Single talker	0	0.547	+2.50	+6.87	+13.84	+23.20
	3	0.381	+2.50	+3.75	+8.53	+16.02
	6	0.305	-1.35	+0.73	+0.52	+8.53
	9	0.212	+1.87	+4.06	+6.04	+7.08
	15	0.175	+1.56	+1.46	+2.91	+4.16
	20	0.187	-0.62	+1.56	+1.66	+3.23

Maxine BNR

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Clean	0.136	0.163 +2.71 [+1.04, +4.80] 12/54/34	0.160 +2.39 [+0.60, +4.55] 20/42/38	0.156 +1.98 [+0.42, +4.22] 16/56/28	0.151 +1.46 [+0.11, +2.80] 14/54/32
Ambient	0.192	0.217 +2.50 [+1.23, +3.98] 16/44/40	0.213 +2.08 [+0.84, +3.44] 20/42/38	0.225 +3.33 [+2.09, +4.76] 15/42/42	0.223 +3.10 [+1.45, +5.15] 17/44/40
Hum	0.160	0.183 +2.29 [+0.64, +3.92] 24/40/37	0.177 +1.72 [+0.24, +3.37] 22/42/36	0.185 +2.45 [+0.79, +4.17] 23/42/36	0.187 +2.68 [+0.73, +4.90] 24/38/38
Broadband noise	0.165	0.200 +3.54 [+1.86, +5.47] 17/44/39	0.189 +2.42 [+1.32, +3.61] 17/45/38	0.200 +3.49 [+2.08, +5.07] 16/40/43	0.193 +2.78 [+1.70, +4.25] 16/44/40

NOISE	BASELINE WER	0.25	0.50	0.75	1.00
Pink noise	0.211	0.244 +3.33 [+1.66, +5.15] 18/40/42	0.251 +4.03 [+2.02, +6.30] 20/37/42	0.245 +3.41 [+1.91, +5.40] 17/40/42	0.236 +2.50 [+1.02, +4.18] 20/39/41
Multi-talker	0.169	0.181 +1.17 [+0.08, +2.38] 22/49/28	0.191 +2.21 [+0.41, +4.37] 24/45/30	0.189 +2.03 [+0.57, +3.61] 22/46/32	0.183 +1.46 [+0.23, +3.03] 20/48/32
Single talker	0.220	0.226 +0.60 [-1.03, +2.16] 26/40/34	0.229 +0.96 [-0.57, +2.51] 24/38/38	0.237 +1.69 [+0.32, +3.16] 22/42/36	0.236 +1.66 [+0.21, +3.10] 22/38/40

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
Clean		0.136	+2.71	+2.39	+1.98	+1.46
Ambient	0	0.554	-0.73	-0.52	+0.94	-0.73
	3	0.343	+6.04	+5.41	+5.10	+5.62
	6	0.270	+1.98	+1.35	+2.50	+2.39
	9	0.201	+1.98	+1.77	+2.50	+1.66
	15	0.157	+2.60	+1.04	+4.37	+3.12
	20	0.140	+3.43	+4.16	+3.95	+5.20
Hum	0	0.202	+1.87	+2.71	+2.60	+2.19
	3	0.179	+0.83	+2.71	+2.19	+2.50
	6	0.162	+1.56	+3.33	+4.16	+3.54
	9	0.159	+1.04	+1.25	+0.94	+0.83
	15	0.163	+2.91	+1.66	+3.43	+2.39
	20	0.156	+3.64	+0.62	+1.25	+3.95
Broadband noise	0	0.287	+1.35	-0.62	-0.31	+1.04
	3	0.260	-1.04	-0.31	-0.42	+2.19
	6	0.200	+3.43	+3.23	+3.95	+2.91
	9	0.171	+2.39	+1.25	+1.77	+1.87
	15	0.149	+3.85	+1.66	+3.54	+2.29
	20	0.140	+4.47	+3.54	+4.68	+4.06
Pink noise	0	0.504	+2.19	+1.66	+1.87	+0.52
	3	0.370	+0.52	+0.83	+1.04	+1.14
	6	0.294	+2.39	+2.39	+3.02	+1.98
	9	0.233	+3.33	+3.54	+4.27	+3.54
	15	0.165	+2.08	+7.70	+3.12	+2.08
	20	0.151	+5.52	+2.50	+3.23	+2.39
Multi-talker	0	0.402	+1.77	+1.66	+3.23	+1.98
	3	0.289	+1.46	+1.66	+2.71	+2.08
	6	0.204	+0.94	+1.14	+1.56	+2.39
	9	0.169	+1.46	+0.94	+1.56	+1.77
	15	0.157	+0.73	+2.71	+1.46	+0.52
	20	0.146	+1.56	+4.06	+3.54	+1.14
Single talker	0	0.547	+2.71	+4.79	+4.79	+4.58
	3	0.381	+2.81	+3.43	+4.47	+3.75
	6	0.305	-1.46	-0.62	+0.21	+0.94

NOISE	SNR	BASELINE WER	0.25	0.50	0.75	1.00
	9	0.212	+1.66	+2.08	+3.54	+1.77
	15	0.175	+2.08	+2.60	+2.19	+2.81
	20	0.187	+0.10	-0.21	+0.83	+1.14

Krisp NC

Pooled, 6 to 20 dB. Each cell: WER / Δ pp / 95% CI / win / neutral / hurt %.

NOISE	BASELINE WER	NSL 50	NSL 100
Clean	0.136	0.164 +2.81 [+0.00, +7.94] 16/52/32	0.159 +2.29 [-0.52, +6.78] 18/56/26
Ambient	0.192	0.226 +3.36 [-0.38, +7.92] 33/40/26	0.270 +7.78 [+3.59, +12.21] 22/36/42
Hum	0.160	0.206 +4.55 [+1.08, +8.20] 25/47/28	0.177 +1.69 [-1.78, +5.41] 34/40/26
Broadband noise	0.165	0.210 +4.50 [+1.69, +7.88] 20/48/32	0.226 +6.14 [+3.23, +9.42] 20/34/46
Pink noise	0.211	0.215 +0.36 [-4.37, +5.85] 44/31/26	0.226 +1.53 [-1.77, +5.80] 40/28/32
Multi-talker	0.169	0.214 +4.47 [+1.11, +8.51] 25/46/30	0.241 +7.26 [+2.62, +12.74] 26/38/36
Single talker	0.220	0.220 +0.05 [-1.51, +1.69] 30/41/30	0.246 +2.65 [+0.87, +4.48] 27/35/38

By SNR. Δ WER (pp) vs baseline; green reduces, red raises.

NOISE	SNR	BASELINE WER	NSL 50	NSL 100
Clean		0.136	+2.81	+2.29
Ambient	0	0.554	-12.17	+2.50
	3	0.343	-1.25	+3.54
	6	0.270	-1.35	+2.71
	9	0.201	-2.19	+2.29
	15	0.157	+7.80	+11.86
	20	0.140	+9.16	+14.26
Hum	0	0.202	-1.46	-4.16
	3	0.179	-1.77	-3.12
	6	0.162	-0.42	-0.52
	9	0.159	+1.46	-1.14
	15	0.163	+9.89	+4.89
	20	0.156	+7.28	+3.54
Broadband noise	0	0.287	-5.62	+0.42

NOISE	SNR	BASELINE WER	NSL 50	NSL 100
	3	0.260	-6.14	-0.73
	6	0.200	-0.83	+3.43
	9	0.171	+0.10	+3.43
	15	0.149	+8.12	+10.82
	20	0.140	+10.61	+6.87
Pink noise	0	0.504	-15.30	-12.80
	3	0.370	-8.32	-4.27
	6	0.294	-8.74	-3.43
	9	0.233	-6.97	-3.23
	15	0.165	+10.61	+7.39
	20	0.151	+6.56	+5.41
Multi-talker	0	0.402	-2.91	+3.02
	3	0.289	-2.29	+1.25
	6	0.204	+2.50	+5.93
	9	0.169	+1.35	+5.20
	15	0.157	+4.68	+7.91
	20	0.146	+9.37	+9.99
Single talker	0	0.547	-4.58	+1.98
	3	0.381	-0.94	+6.87
	6	0.305	-1.35	+3.23
	9	0.212	+0.21	+6.35
	15	0.175	+1.46	+1.25
	20	0.187	-0.10	-0.21

A.3 BASELINE WER (NO-NC)

CONDITION	ENGLISH	HINDI
Clean	0.033	0.136
Ambient	0.069	0.192
Hum	0.035	0.160
Broadband noise	0.044	0.165
Pink noise	0.094	0.211
Multi-talker	0.188	0.169
Single talker	0.229	0.220