

SI Appendix

Table S1. Summary Statistics for TRF peak amplitude comparisons between PRE vs POST, PRE vs CLEAN, and POST vs CLEAN and for both early and late peak (paired t values, p values, and Cohen's d for effect size) in the main study. Differences between CLEAN and either PRE or POST are significant except for phoneme onset early peak. Differences between PRE and POST are significant but only for word onset responses.

		Envelope	Envelope Onset	Phoneme Onset	Word onset
PRE vs POST	Early peak	$t_{24} = 0.48, p = 0.63$; Cohen's d = 0.03	$t_{24} = 0.25, p = 0.80$; d = 0.02	$t_{24} = -0.52, p = 0.61$; d = 0.04	$t_{24} = -2.37, p = \mathbf{0.02}$; d = 0.39
	Late Peak	$t_{24} = -0.37, p = 0.72$; d = 0.04	$t_{24} = 1.09, p = 0.29$; d = 0.21	$t_{24} = -1.03, p = 0.31$; d = 0.29	$t_{24} = -3.87, p < \mathbf{0.001}$; d = 0.87
PRE vs CLEAN	Early peak	$t_{24} = 6.04, p < \mathbf{0.001}$; d = 1.04	$t_{24} = -4.28, p < \mathbf{0.001}$; d = 0.70	$t_{24} = 1.1, p = 0.28$; d = 0.11	$t_{24} = -4.77, p < \mathbf{0.001}$; d = 1.34
	Late Peak	$t_{24} = -2.37, p = \mathbf{0.03}$; d = 0.54	$t_{24} = -3.37, p = \mathbf{0.003}$; d = 0.74	$t_{24} = -3.24, p = \mathbf{0.003}$; d = 0.93	$t_{24} = -5.21, p < \mathbf{0.001}$; d = 1.20
POST vs CLEAN	Early peak	$t_{24} = 6.63, p < \mathbf{0.001}$; d = 1.09	$t_{24} = -4.35, p < \mathbf{0.001}$; d = 0.70	$t_{24} = -1.16, p = 0.25$; d = 0.13	$t_{24} = -4.59, p = \mathbf{0.001}$; d = 1.14
	Late Peak	$t_{24} = -2.29, p = \mathbf{0.03}$; d = 0.5	$t_{24} = -3.32, p = \mathbf{0.003}$; d = 0.83	$t_{24} = -3.89, p < \mathbf{0.001}$; d = 0.53	$t_{24} = -2.18, p = \mathbf{0.04}$; d = 0.40

Table S2. Summary Statistics for TRF peak amplitude comparisons between PRE1 vs PRE2, PRE1 vs CLEAN, and PRE2 vs CLEAN and for both early and late peak (paired t values, p values, and Cohen's d for effect size) in the control study. Differences between CLEAN and either PRE1 or PRE2 are significant for early peaks except for phoneme onset. Almost all differences between PRE1 and PRE2 are not significant.

		Envelope	Envelope Onset	Phoneme Onset	Word onset
PRE1 vs PRE2	Early peak	$t_{11} = -4.23, p = \mathbf{0.001}$; Cohen's d = 0.31	$t_{11} = 0.55, p = 0.59$; d = 0.09	$t_{11} = -1.76, p = 0.11$; d = 0.23	$t_{11} = 1.01, p = 0.34$; d = 0.22
	Late Peak	$t_{11} = 0.08, p = 0.94$; d = 0.02	$t_{11} = 1.79, p = 0.10$; d = 0.49	$t_{11} = 0.66, p = 0.52$; d = 0.14	$t_{11} = 0.35, p = 0.74$; d = 0.06
PRE1 vs CLEAN	Early peak	$t_{11} = 4.41, p = \mathbf{0.001}$; d = 1.01	$t_{11} = -3.90, p = \mathbf{0.002}$; d = 0.87	$t_{11} = -0.19, p = 0.85$; d = 0.02	$t_{11} = -4.98, p < \mathbf{0.001}$; d = 1.63
	Late Peak	$t_{11} = -2.32, p = 0.04$; d = 0.83	$t_{11} = -2.31, p = 0.04$; d = 0.84	$t_{11} = -2.84, p = 0.02$; d = 0.63	$t_{11} = -5.11, p = \mathbf{0.009}$; d = 1.76
PRE2 vs CLEAN	Early peak	$t_{11} = 4.77, p < \mathbf{0.001}$; d = 1.27	$t_{11} = -3.43, p = \mathbf{0.006}$; d = 0.96	$t_{11} = 1.39, p = 0.19$; d = 0.21	$t_{81} = -4.71, p < \mathbf{0.001}$; d = 1.84
	Late Peak	$t_{11} = -2.12, p = 0.06$; d = 0.78	$t_{11} = -2.88, p = 0.02$; d = 1.07	$t_{11} = -2.50, p = 0.03$; d = 0.80	$t_{11} = -5.26, p < \mathbf{0.001}$; d = 1.86

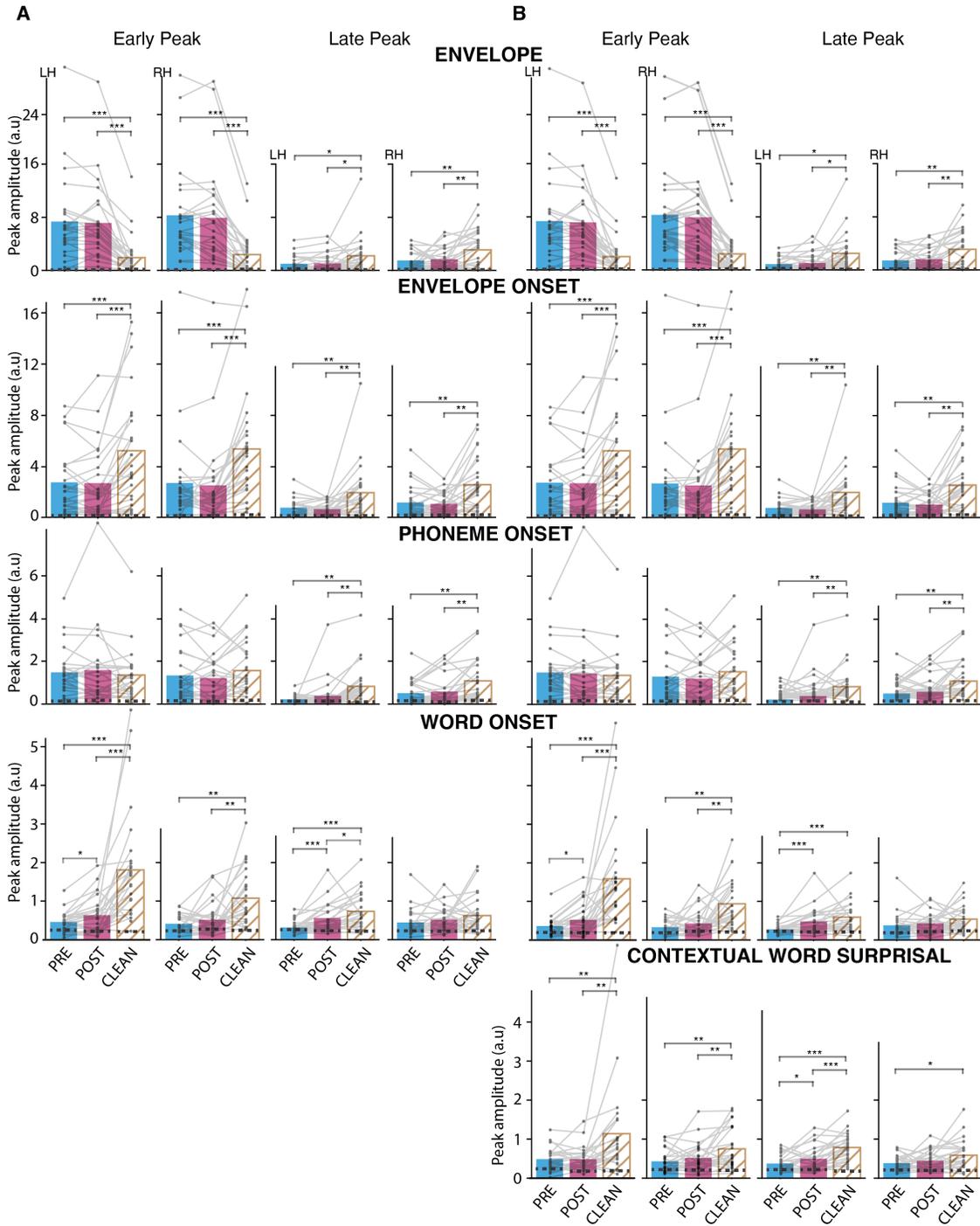


Figure S1. TRF peak amplitude comparison by speech condition for envelope, envelope onset, phoneme onset, word onset and contextual word surprisal, for early and late peaks, and by hemisphere (A) for the same TRF model as in Figure 2 and (B) for another model including contextual word surprisal. The dashed lines within the bars represent the noise floor. Other details as in Figure 2. Similar to the main model, almost all differences between CLEAN and either PRE or POST are significant, except for phoneme onset early peak. Differences between PRE and POST are significant only for word onset responses and contextual word surprisal late responses

in the left hemisphere. Adding contextual word surprisal into the model does not substantially change significant intelligibility effects observed in word onset responses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

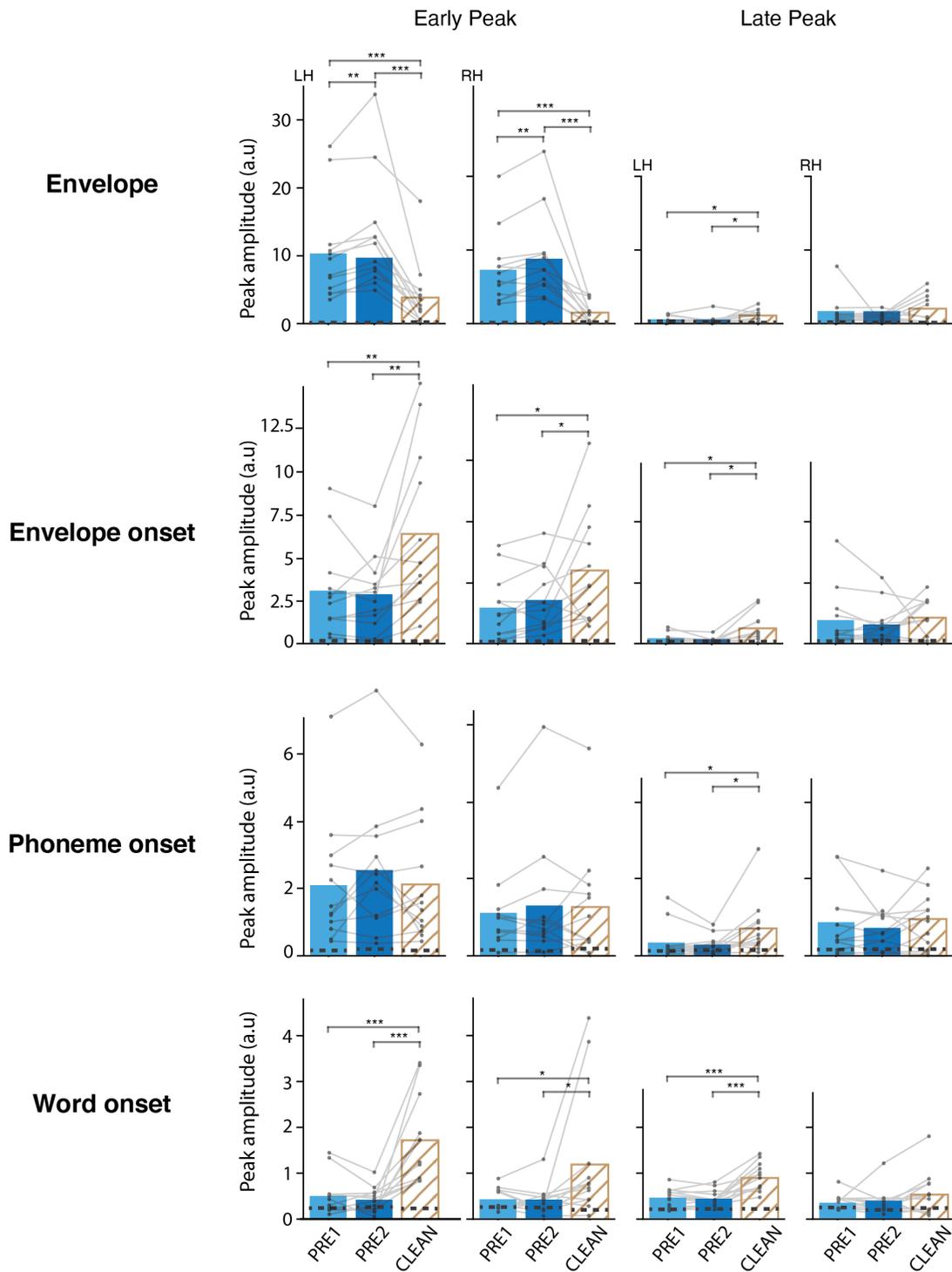


Figure S2. Control study TRF peak amplitude comparison by speech condition for envelope, envelope onset, phoneme onset, and word onset, for early and late peaks, and by hemisphere. Other details as in Figure 2. Differences between PRE1 and PRE2 are significant only for early envelope responses in both hemispheres. Differences between vocoded and CLEAN are similar to those observed in the main study.

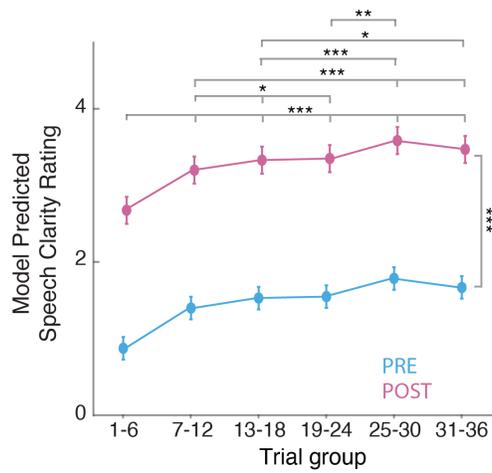


Figure S3. Speech clarity rating change over trials. Trials were combined using a tumbling window of 6 trials. Linear Mixed Effects Model (clarity rating ~ 1 + condition + trial group (1+condition|Subject)) predicted clarity rating \pm SE change over trial group. As no significant interaction was found between condition and trial group, the model's trends over trials for both PRE and POST are parallel. Overall, speech clarity improves over trials for both PRE and POST vocoded speech (trials 1-6 as the reference, trials 7-12 = 0.53, SE = 0.06, $p < 0.001$, trials 13-18 = 0.66, SE = 0.06, $p < 0.001$, trials 19-24 = 0.68, SE = 0.06, $p < 0.001$, trials 25-30 = 0.91, SE = 0.06, $p < 0.001$, trials 31-36 = 0.79, SE = 0.06, $p < 0.001$) and then level off (trials 19-24 to trials 25-30 = 0.15, SE = 0.06, $p = 0.33$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$