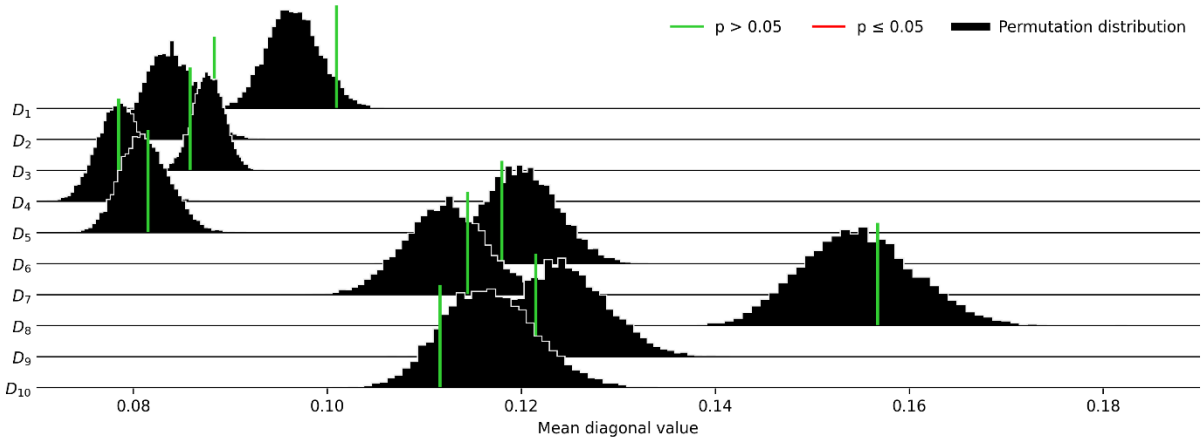
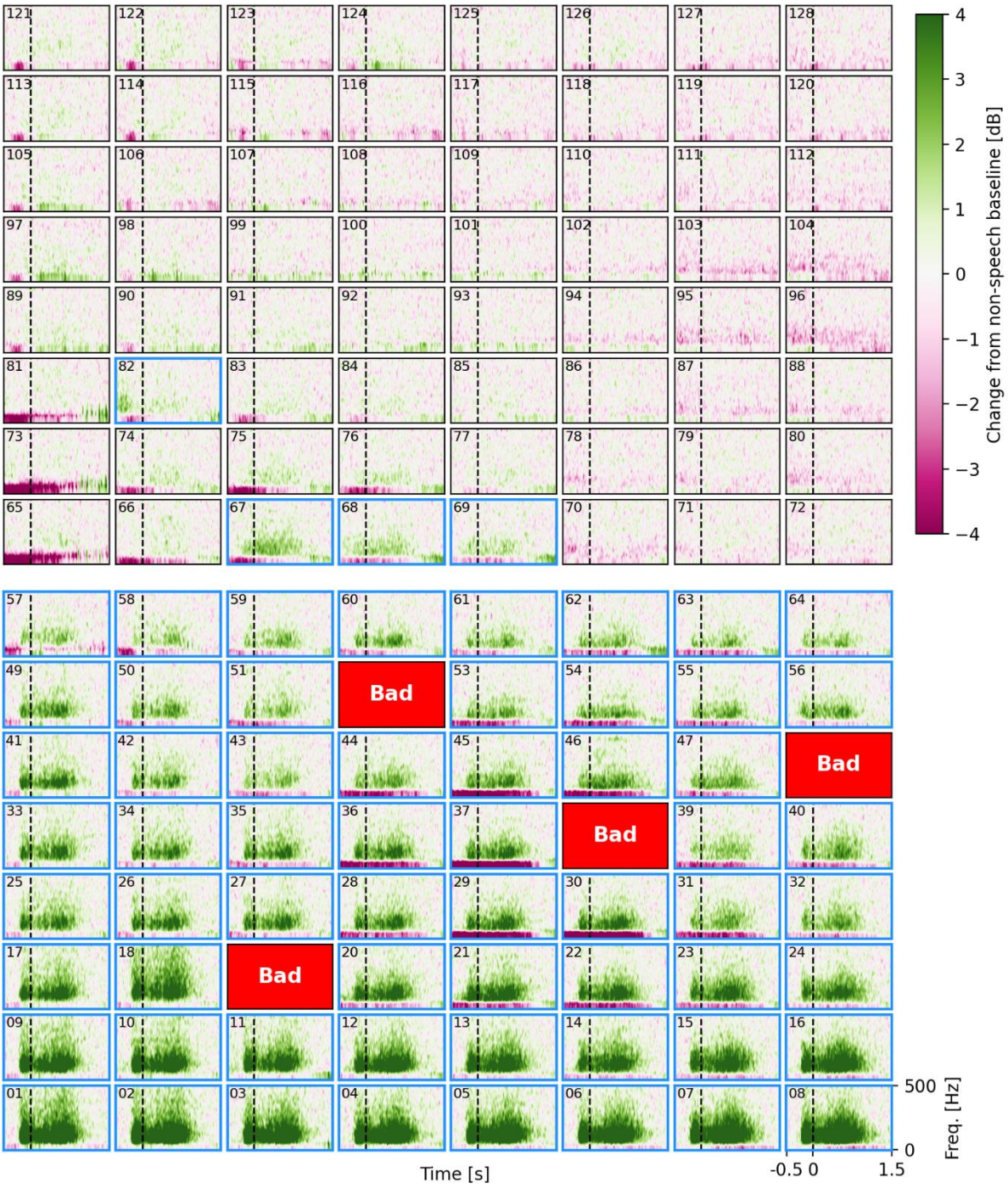
Supplementary Materials

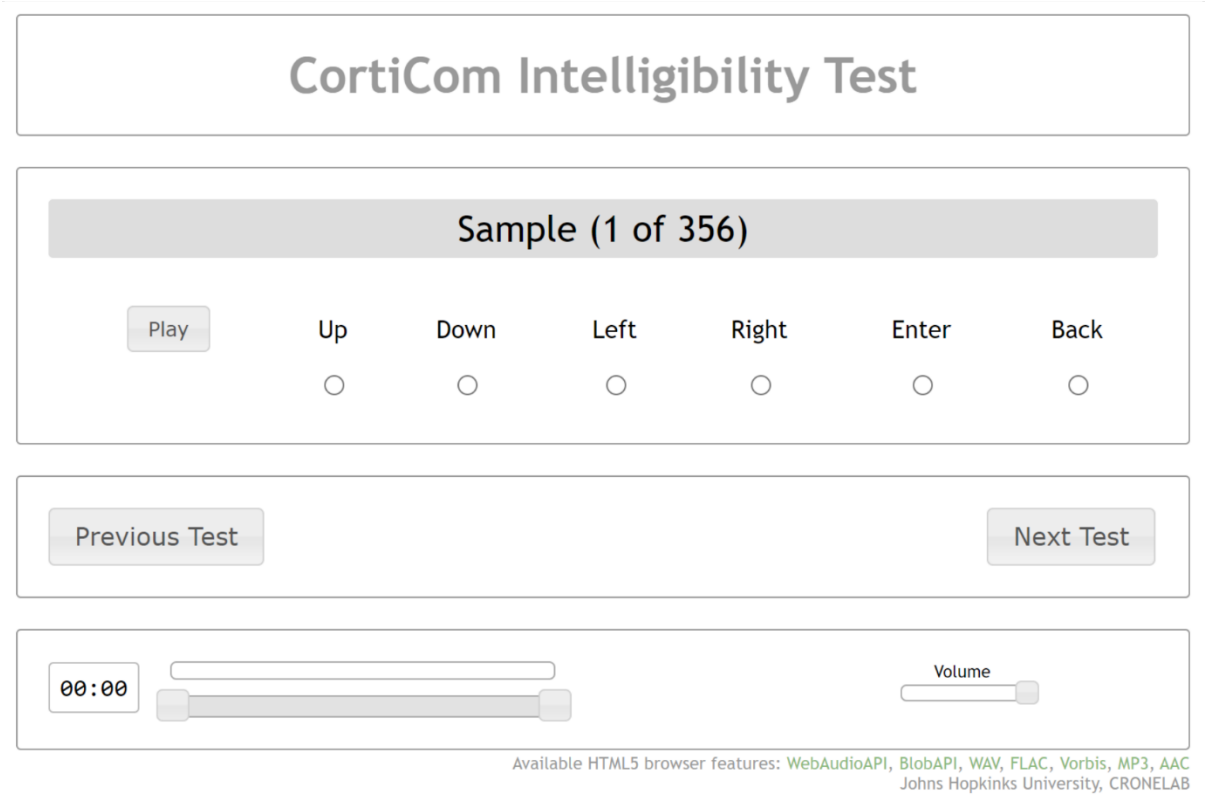


Supplementary Figure 1 | Acoustic contamination report. We checked all recordings in our training, validation and test sets for acoustic contamination using the method proposed by Roussel et al. (Roussel, 2020). For each day, the histogram represents mean diagonal values of the distribution of permutated contamination matrices (N=10,000). Colored bars indicate the position of the contamination index of neural recordings, while the statistical criterion for rejecting the null hypothesis is indicated in green ($p > 0.05$) or in red ($p \leq 0.05$) if acoustic contamination is present. After correcting Channel 46 for D_5, D_6, and D_7 we did not detect any acoustic contamination. Recordings of the closed-loop sessions where not checked for acoustic contamination since they were completely withheld from network training procedures, and therefore cannot leak into model training and lead to the Clever Hans phenomenon (Lapuschkin, 2019).



Supplementary Figure 2 | Power spectral analysis for identifying speech-related channels. Each plot shows channel specific normalized power spectral density in the frequency range of 0 - 500 Hz averaged across all keywords for a particular recording session in the training set (60 trials). Dashed lines indicate speech onset times, while the x-axis shows time-aligned power spectral density in the range of -0.5 s up

to 1.5 s. Bad channels were excluded based on visual inspection. Channels marked with blue borders represent the top-64 channels whose lower bound of a 95% confidence interval most often exceed baseline high-gamma activity (see Supplementary Data for upper and lower bounds, training set only) and were selected as relevant channels for the speech synthesis task.



Supplementary Figure 3 | Web-Interface for conducting the listening test. All synthesized samples are presented in a randomized sequence. Human listeners can use the Play button to play the sample and subsequently select one of the radio buttons to make their choice or play the sample again. After making a choice they can move on to the next sample. Human listeners can go back to the previous sample if they accidentally make a wrong decision, but not beyond it.

Nr.	1	2	3	4	5	6	7	8	9	10	11	12
CV	ZOO	VOO	HEE	BAA	GEE	THEE	HAA	JEE	LAA	TOO	YAA	MOO

Supplementary Table 1 | Syllable stimuli from the syllable repetition task. Based on Bouchard et al. (Bouchard, 2013) we chose 12 consonant-vowel (CV) syllable stimuli to acquire high-gamma activity of articulatory representations for a stable baseline across days.