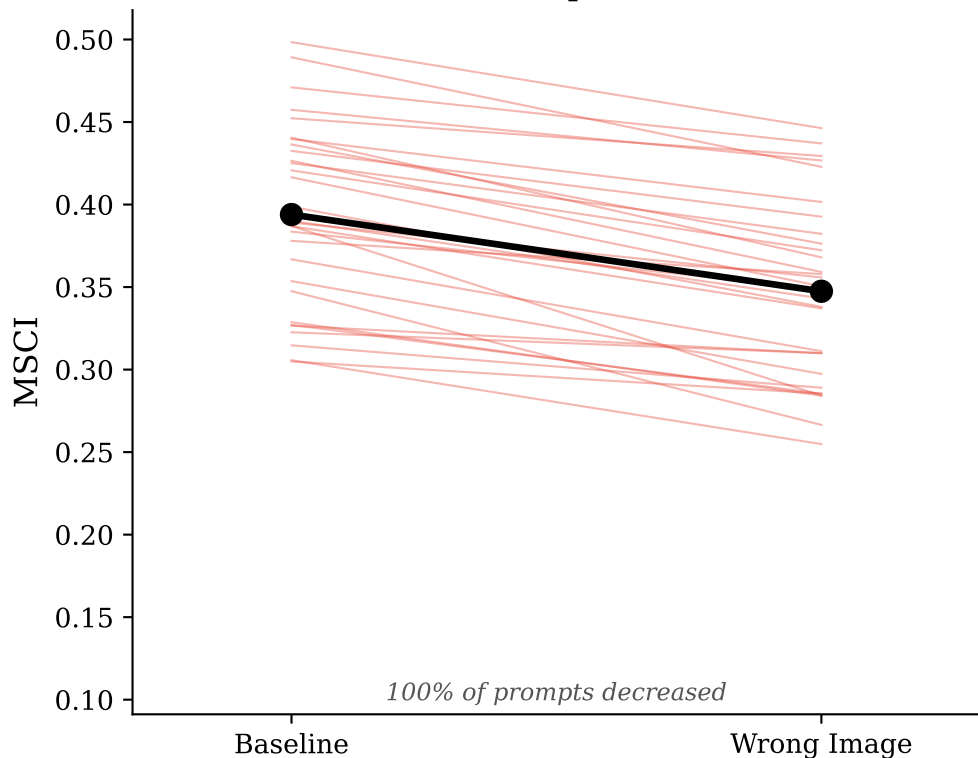


RQ1: Per-Prompt MSCI Change Under Perturbation

Baseline \rightarrow Wrong Image
($d = 2.27$, $p < .001$)



Baseline \rightarrow Wrong Audio
($d = 3.64$, $p < .001$)

