

RQ3: MSCI vs Human Coherence Judgment

Spearman's $\rho = 0.379$ ($p = 0.039$)
Kendall's $\tau = 0.293$ ($p = 0.038$)
 $N = 30$, $k = 3$ raters

Human Coherence Rating (median across raters)

5
4
3
2
1

MSCI Score

- Baseline
- Wrong Image
- Wrong Audio

0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50

