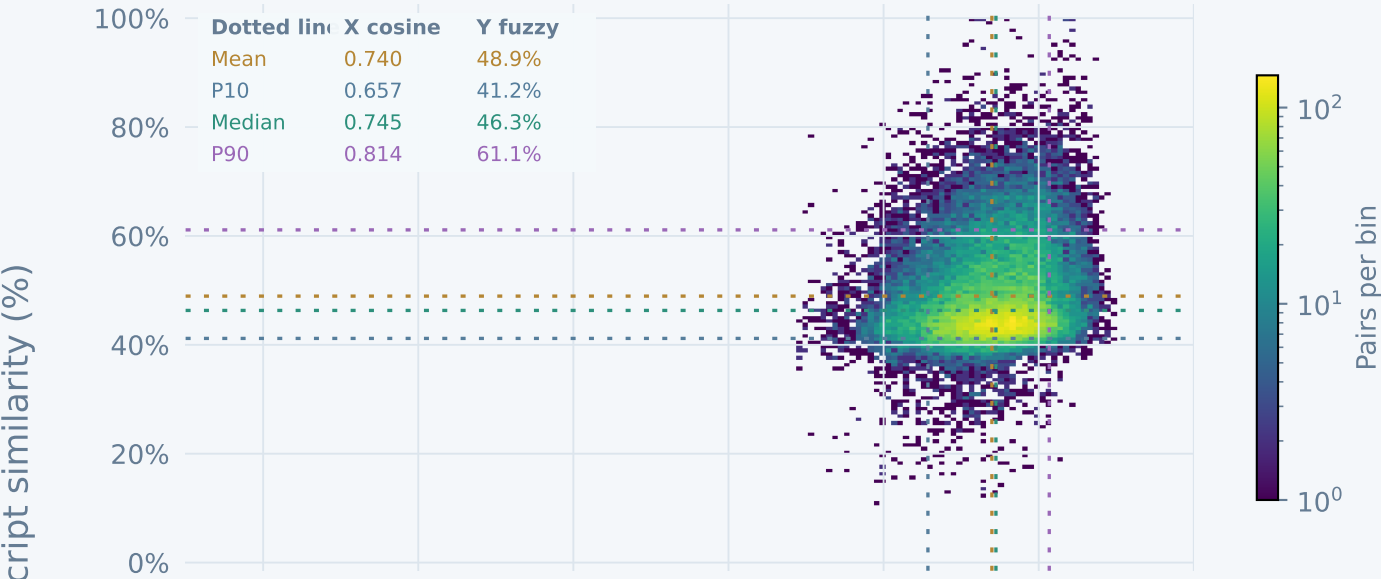


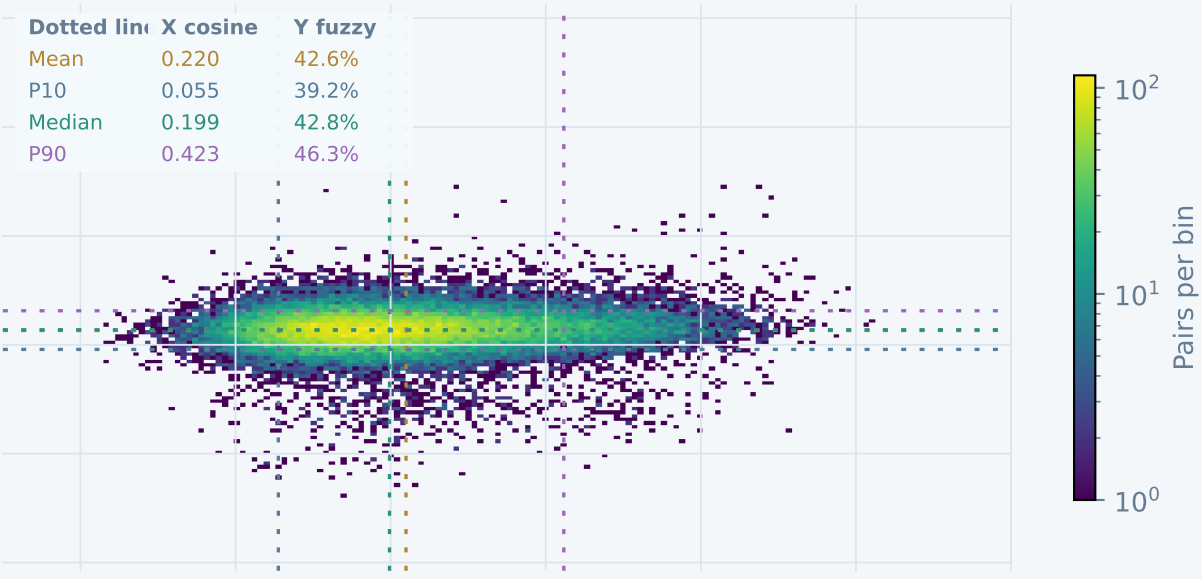
Voice similarity × transcript similarity

2,000 uniform random queries • 838,411 pair measurements • Exact text matches included

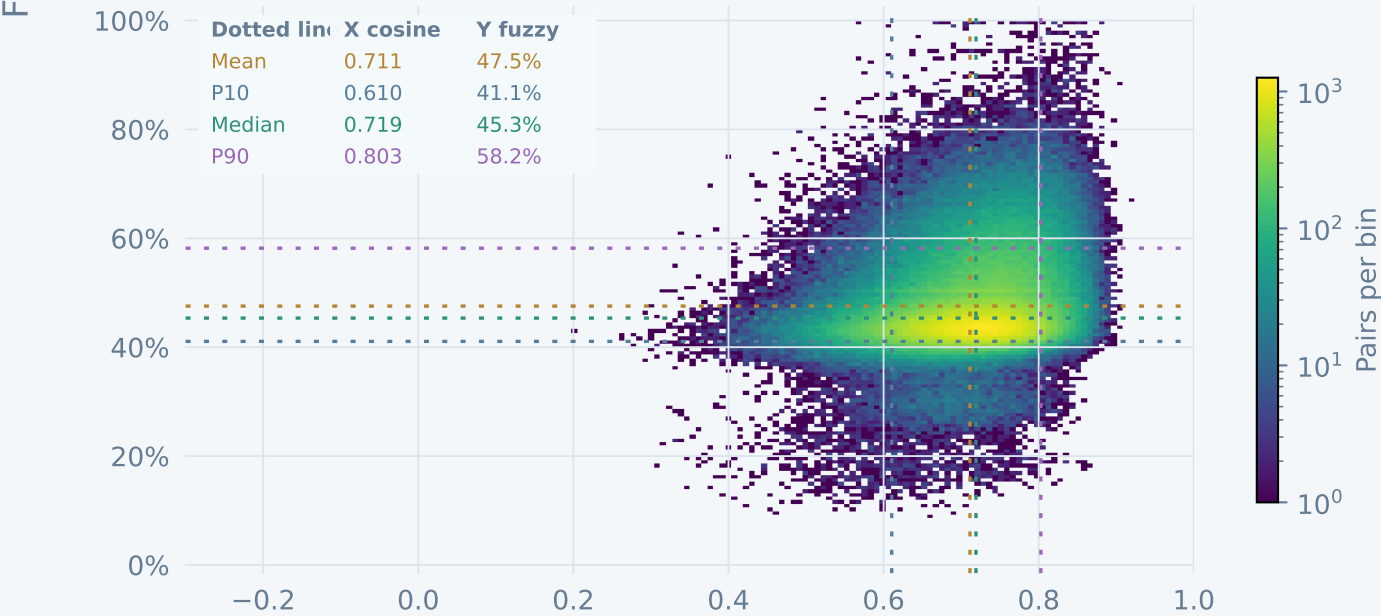
Query → nearest results
40,000 pairs • Spearman $\rho = 0.173$



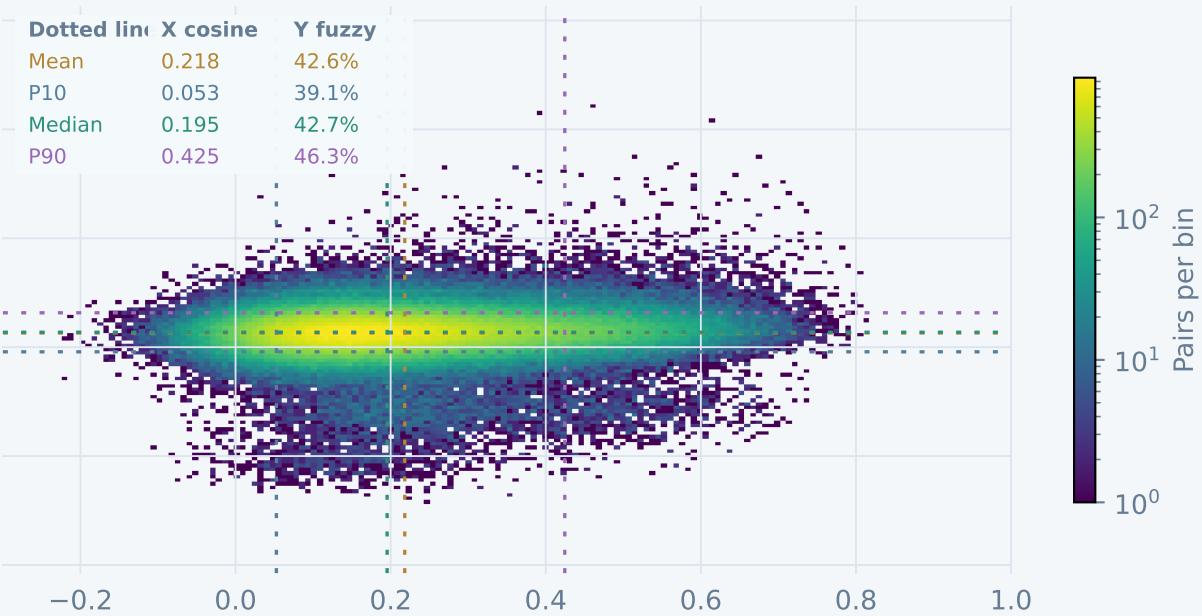
Query → matched random
39,888 pairs • Spearman $\rho = 0.066$



Between nearest results
380,000 pairs • Spearman $\rho = 0.212$



Between matched random
378,523 pairs • Spearman $\rho = 0.060$



Speaker cosine similarity

Finding: retrieved pairs show a modest positive association ($r \approx 0.17$ – 0.22); high speaker similarity often occurs with different wording.
Dotted lines: pair-weighted mean, P10, median (P50) and P90 on each axis. These are marginal summaries, not trends or confidence intervals.
Near-identical scripts occur in pockets. Selection and shared generation inputs can contribute; correlation does not establish identity or causal content leakage.