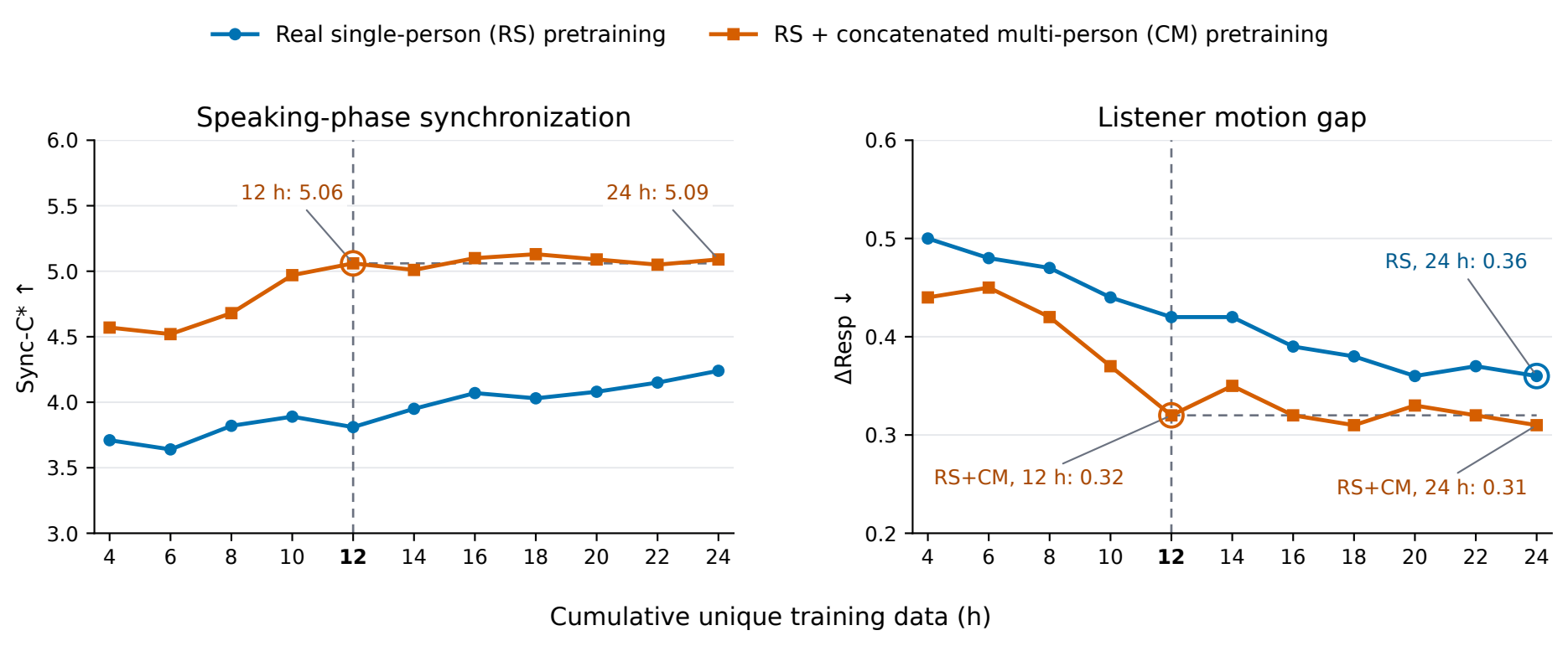

Real single-person (RS) pretraining
RS + concatenated multi-person (CM) pretraining
Speaking-phase synchronization
Sync-C* ↑
12 h: 5.06
24 h: 5.09
Listener motion gap
ΔResp ↓
RS, 24 h: 0.36
RS+CM, 12 h: 0.32
RS+CM, 24 h: 0.31
Cumulative unique training data (h)