Normalized confusion matrix (gpt-4o-mini) action -0.34 0.02 0.00 0.03 0.03 0.01 0.32 0.11 0.01 0.03 0.08 0.01 0.00 aggression -0.03 0.83 0.00 0.00 0.00 0.00 0.05 0.03 0.00 0.01 0.05 0.00 0.00 delivery -0.13 0.02 0.01 0.44 0.02 0.01 0.17 0.06 0.05 0.08 0.00 0.00 0.00 entrance -0.00 0.00 0.00 0.00 0.89 0.04 0.02 0.01 0.00 0.03 0.00 0.02 0.00 exit -0.00 0.00 0.00 0.00 0.17 0.78 0.01 0.03 0.00 0.01 0.00 0.00 0.00 True label interaction -0.05 0.06 0.00 0.00 0.02 0.01 0.78 0.04 0.00 0.00 0.04 0.00 0.00 movement -0.03 0.00 0.00 0.00 0.03 0.18 0.17 0.57 0.00 0.02 0.00 0.01 0.00 music -0.01 0.00 0.11 0.03 0.01 0.01 0.02 0.02 0.69 0.03 0.03 0.03 0.01 narration -0.04 0.05 0.00 0.00 0.02 0.01 0.05 0.02 0.03 0.72 0.04 0.03 0.00 object -0.06 0.03 0.00 0.07 0.01 0.00 0.09 0.03 0.01 0.03 0.66 0.00 0.00 setting -0.02 0.00 0.01 0.00 0.03 0.02 0.02 0.09 0.00 0.04 0.01 0.77 0.00 toward -0.00 0.00 0.00 0.02 0.16 0.00 0.20 0.01 0.00 0.00 0.00 0.07 0.52 actionession movement interaction narration Music Predicted label