Normalized confusion matrix (gpt-4o-mini) action -0.38 0.03 0.00 0.02 0.05 0.03 0.19 0.05 0.00 0.07 0.14 0.01 0.03 aggression -0.14 0.73 0.00 0.00 0.00 0.05 0.05 0.00 0.05 0.00 0.00 0.00 delivery -0.17 0.03 0.00 0.47 0.02 0.00 0.09 0.03 0.05 0.11 0.00 0.00 0.03 entrance -0.03 0.00 0.00 0.00 0.84 0.03 0.00 0.00 0.03 0.03 0.00 0.03 0.03 True label movement -0.22 0.00 0.00 0.00 0.03 0.14 0.11 0.39 0.00 0.08 0.00 0.00 0.03 music -0.02 0.00 0.00 0.08 0.01 0.00 0.00 0.00 0.80 0.06 0.02 0.02 0.00 narration -0.08 0.03 0.00 0.08 0.03 0.00 0.11 0.00 0.00 0.56 0.03 0.08 0.00 object -0.19 0.00 0.00 0.02 0.00 0.00 0.19 0.00 0.00 0.03 0.55 0.00 0.02 setting -0.05 0.00 0.02 0.00 0.00 0.04 0.02 0.11 0.00 0.05 0.04 0.67 0.02 action action movement interaction narration MUSIC Predicted label