Normalized confusion matrix (gpt-4o-mini) action -0.44 0.03 0.01 0.03 0.03 0.01 0.23 0.14 0.01 0.03 0.03 0.02 0.00 aggression -0.05 0.86 0.00 0.00 0.00 0.00 0.05 0.00 0.05 0.00 0.00 0.00 delivery -0.16 0.02 0.02 0.41 0.00 0.02 0.19 0.03 0.09 0.08 0.00 0.00 0.00 entrance -0.00 0.00 0.00 0.00 0.89 0.05 0.00 0.00 0.00 0.00 0.00 0.05 0.00 exit -0.00 0.01 0.00 0.00 0.14 0.79 0.04 0.01 0.00 0.00 0.00 0.00 0.00 True label interaction -0.00 0.16 0.00 0.00 0.00 0.00 <mark>0.84</mark> 0.00 0.00 0.00 0.00 0.00 0.00 movement -0.06 0.00 0.00 0.00 0.03 0.17 0.19 0.56 0.00 0.00 0.00 0.00 0.00 music -0.01 0.01 0.04 0.02 0.02 0.01 0.02 0.00 0.78 0.05 0.02 0.03 0.01 narration -0.25 0.00 0.00 0.00 0.00 0.00 0.19 0.03 0.08 0.36 0.00 0.08 0.00 object -0.27 0.02 0.00 0.29 0.00 0.00 0.19 0.00 0.00 0.02 0.21 0.00 0.00 setting -0.02 0.00 0.02 0.00 0.02 0.04 0.02 0.16 0.00 0.02 0.00 0.72 0.00 toward -0.01 0.00 0.01 0.01 0.05 0.00 0.07 0.01 0.00 0.00 0.00 0.07 0.75 action interaction narration MUSIC Predicted label