## Normalized confusion matrix (gpt-4o-mini) action -0.41 0.02 0.01 0.02 0.08 0.01 0.21 0.12 0.02 0.02 0.07 0.03 0.00 delivery -0.15 0.03 0.00 0.28 0.05 0.01 0.15 0.05 0.20 0.06 0.00 0.01 0.00 entrance -0.00 0.00 0.01 0.00 0.91 0.04 0.00 0.01 0.01 0.01 0.00 0.02 0.00 exit -0.09 0.00 0.00 0.00 0.17 0.66 0.01 0.05 0.00 0.00 0.00 0.00 0.00 True label interaction -0.07 0.12 0.00 0.00 0.02 0.00 <mark>0.74</mark> 0.02 0.00 0.00 0.04 0.00 0.00 movement -0.03 0.00 0.00 0.00 0.06 0.14 0.17 0.57 0.01 0.01 0.00 0.01 0.00 music -0.02 0.01 0.26 0.03 0.04 0.00 0.02 0.02 0.40 0.08 0.04 0.05 0.04 narration -0.11 0.06 0.00 0.01 0.02 0.01 0.16 0.05 0.07 0.41 0.03 0.09 0.00 object -0.43 0.04 0.00 0.13 0.01 0.00 0.12 0.01 0.01 0.01 0.20 0.02 0.00 setting -0.02 0.00 0.01 0.00 0.04 0.02 0.01 0.10 0.01 0.01 0.00 0.79 0.00 toward -0.00 0.00 0.06 0.02 0.08 0.00 0.10 0.01 0.00 0.01 0.00 0.06 0.63 action adjoestion interaction narration MUSIC

Predicted label