Normalized confusion matrix (mistral-small-latest) action -0.62 0.01 0.00 0.01 0.01 0.01 0.06 0.07 0.00 0.01 0.15 0.05 0.00 aggression -0.14 0.73 0.00 0.00 0.00 0.05 0.00 0.00 0.05 0.05 0.05 0.00 0.00 delivery -0.22 0.06 0.00 0.38 0.00 0.00 0.08 0.05 0.05 0.06 0.02 0.09 0.00 entrance -0.03 0.00 0.00 0.00 0.66 0.03 0.00 0.05 0.03 0.00 0.00 0.18 0.03 exit -0.07 0.00 0.00 0.00 0.04 0.53 0.01 0.32 0.00 0.00 0.01 0.01 0.00 True label interaction -0.06 0.13 0.00 0.00 0.00 0.00 <mark>0.77</mark> 0.00 0.00 0.00 0.03 0.00 0.00 movement -0.14 0.00 0.00 0.00 0.03 0.14 0.11 0.56 0.00 0.00 0.03 0.00 0.00 music +0.03 0.01 0.00 0.00 0.01 0.00 0.02 0.00 0.62 0.05 0.03 0.17 0.05 narration -0.11 0.08 0.00 0.00 0.00 0.00 0.14 0.08 0.06 0.28 0.08 0.17 0.00 object -0.10 0.00 0.00 0.00 0.00 0.00 0.15 0.02 0.00 0.00 0.73 0.02 0.00 setting -0.04 0.00 0.00 0.00 0.00 0.02 0.00 0.16 0.00 0.00 0.00 0.79 0.00 action movement interaction narration Music Predicted label