Normalized confusion matrix (mistral-small-latest) action -0.22 0.02 0.01 0.04 0.03 0.02 0.17 0.13 0.02 0.06 0.17 0.03 0.07 aggression -0.09 0.39 0.04 0.01 0.00 0.01 0.19 0.09 0.00 0.07 0.07 0.03 0.01 delivery -0.09 0.03 0.03 0.37 0.02 0.02 0.08 0.03 0.07 0.15 0.01 0.02 0.08 entrance -0.03 0.01 0.00 0.02 0.66 0.04 0.01 0.05 0.05 0.09 0.02 0.03 0.00 exit $+0.02 \ 0.01 \ 0.01 \ 0.00 \ 0.09 \ 0.67 \ 0.01 \ 0.07 \ 0.02 \ 0.03 \ 0.02 \ 0.01 \ 0.02$ True label interaction -0.07 0.03 0.00 0.01 0.01 0.01 <mark>0.66</mark> 0.05 0.01 0.02 0.14 0.00 0.00 movement -0.08 0.00 0.01 0.03 0.08 0.09 0.12 0.47 0.02 0.04 0.03 0.03 0.03 music +0.02 0.01 0.01 0.03 0.02 0.01 0.02 0.01 0.73 0.05 0.05 0.03 0.01 narration -0.09 0.04 0.01 0.02 0.01 0.03 0.05 0.06 0.03 0.54 0.05 0.04 0.03 object -0.06 0.02 0.01 0.01 0.01 0.02 0.13 0.02 0.01 0.03 0.62 0.02 0.01 setting -0.02 0.02 0.02 0.01 0.03 0.01 0.03 0.10 0.02 0.03 0.04 0.66 0.02 toward -0.01 0.02 0.02 0.02 0.01 0.01 0.03 0.01 0.01 0.02 0.01 0.02 0.82 action adjoestion movement interaction narration aparte MUSIC delivery Predicted label