Normalized confusion matrix (gpt-4o) action -0.38 0.01 0.00 0.01 0.02 0.01 0.30 0.09 0.01 0.02 0.13 0.01 0.01 aggression -0.04 0.75 0.00 0.00 0.00 0.00 0.07 0.05 0.00 0.01 0.08 0.00 0.00 delivery -0.22 0.01 0.00 0.54 0.04 0.01 0.08 0.01 0.02 0.06 0.00 0.00 0.00 entrance -0.01 0.00 0.00 0.00 0.88 0.05 0.02 0.02 0.01 0.02 0.00 0.00 0.00 exit +0.00 0.00 0.00 0.00 0.15 0.79 0.00 0.04 0.00 0.00 0.00 0.00 0.00 True label interaction -0.03 0.03 0.00 0.00 0.02 0.01 0.81 0.03 0.00 0.01 0.06 0.00 0.00 movement -0.03 0.00 0.00 0.00 0.05 0.18 0.07 0.64 0.00 0.01 0.01 0.00 0.03 music -0.01 0.00 0.00 0.02 0.01 0.00 0.00 0.01 0.89 0.01 0.02 0.01 0.01 narration $\pm 0.03 \ 0.06 \ 0.00 \ 0.00 \ 0.03 \ 0.02 \ 0.03 \ 0.04 \ 0.02 \ 0.68 \ 0.04 \ 0.05 \ 0.00$ setting -0.02 0.00 0.00 0.01 0.02 0.02 0.02 0.09 0.01 0.02 0.01 0.81 0.00 toward $+0.01\ 0.00\ 0.00\ 0.02\ 0.01\ 0.00\ 0.03\ 0.00\ 0.00\ 0.00\ 0.00\ 0.06\ 0.85$ action adjoestion movement interaction narration Music Predicted label