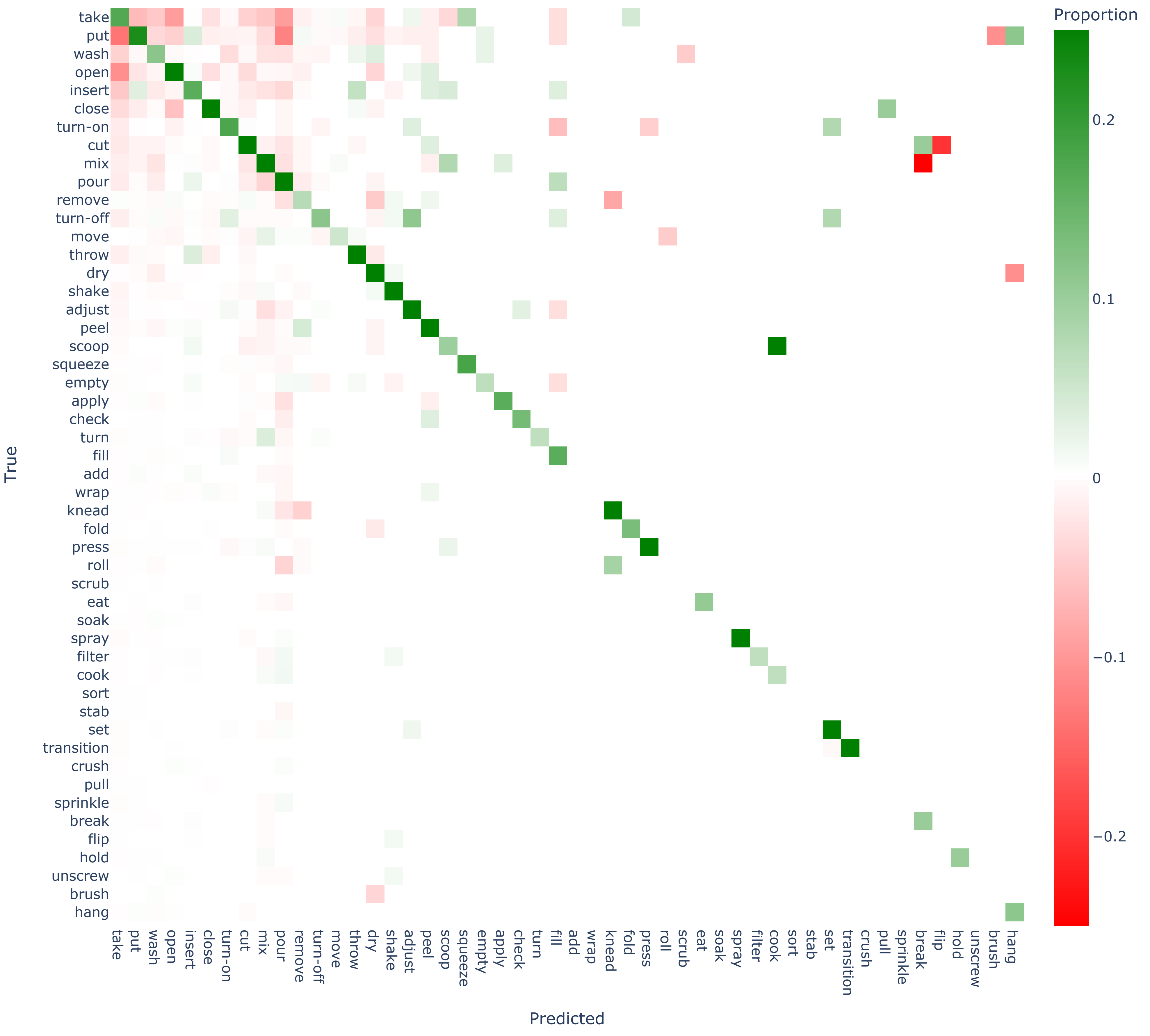
Difference in confusion matrix for the top 50 verb classes
between **multi-modal ensemble** and AudioSlowFast


True
take
put
wash
open
insert
close
turn-on
cut
mix
pour
remove
turn-off
move
throw
dry
shake
adjust
peel
scoop
squeeze
empty
apply
check
turn
fill
add
wrap
knead
fold
press
roll
scrub
eat
soak
spray
filter
cook
sort
stab
set
transition
crush
pull
sprinkle
break
flip
hold
unscrew
brush
hang
Predicted
take
put
wash
open
insert
close
turn-on
cut
mix
pour
remove
turn-off
move
throw
dry
shake
adjust
peel
scoop
squeeze
empty
apply
check
turn
fill
add
wrap
knead
fold
press
roll
scrub
eat
soak
spray
filter
cook
sort
stab
set
transition
crush
pull
sprinkle
break
flip
hold
unscrew
brush
hang
Proportion