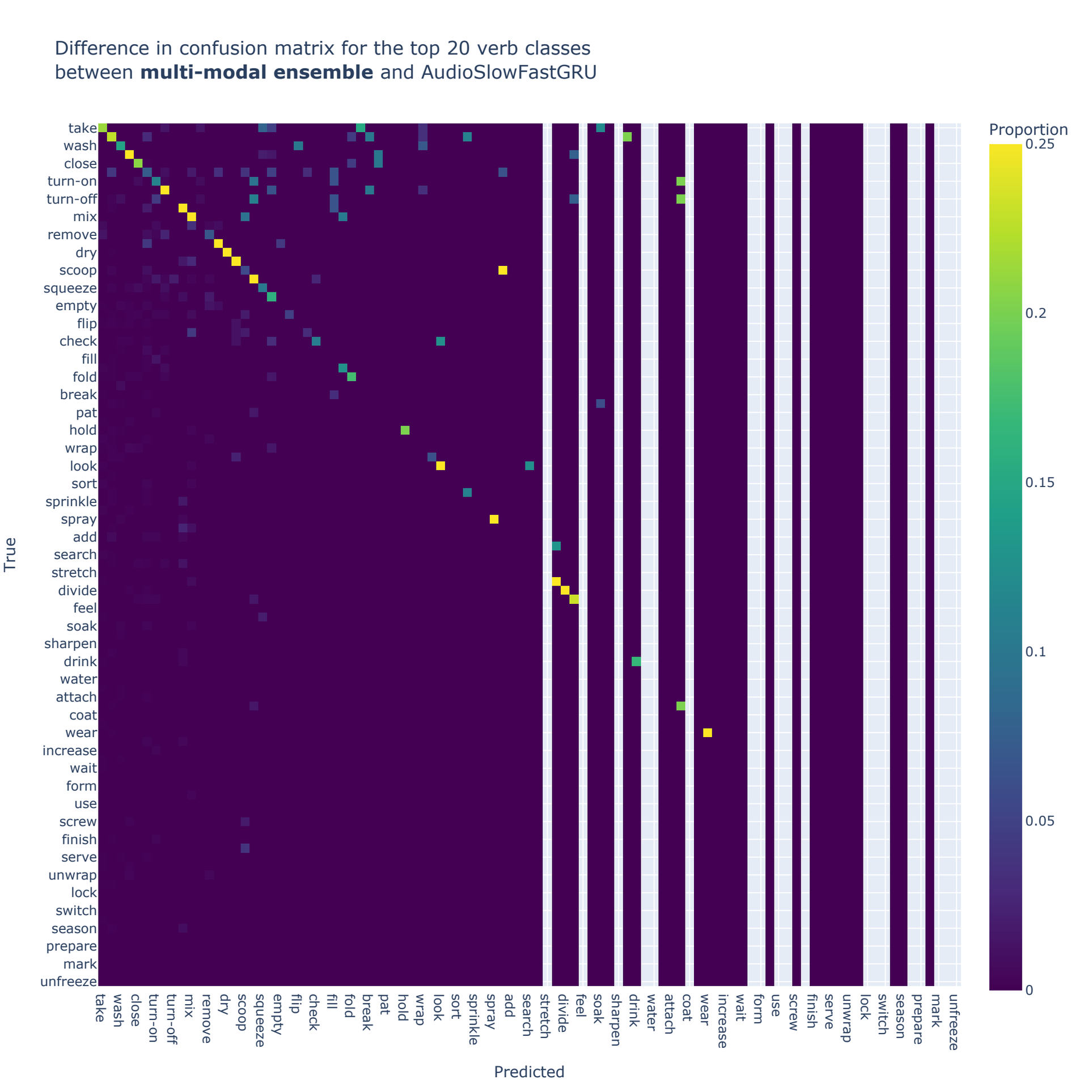
Difference in confusion matrix for the top 20 verb classes
between **multi-modal ensemble** and AudioSlowFastGRU
True
Predicted
Proportion
0.25
0.2
0.15
0.1
0.05
0
take
wash
close
turn-on
turn-off
mix
remove
dry
scoop
squeeze
empty
flip
check
fill
fold
break
pat
hold
wrap
look
sort
sprinkle
spray
add
search
stretch
divide
feel
soak
sharpen
drink
water
attach
coat
wear
increase
wait
form
use
screw
finish
serve
unwrap
lock
switch
season
prepare
mark
unfreeze