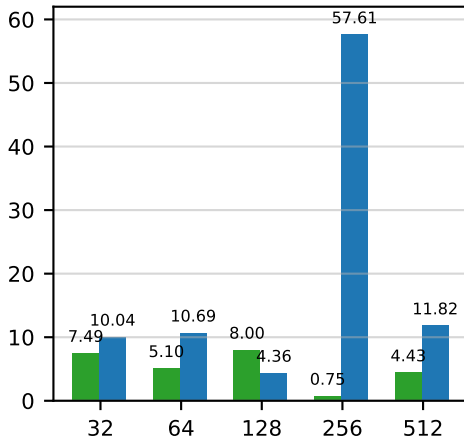
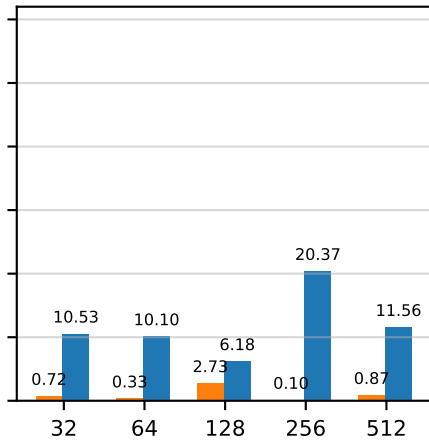


Averaged Importance Weight ( $\times 10^2$ )

Audio Encoder



Visual Encoder



Working memory



Vision



Audio