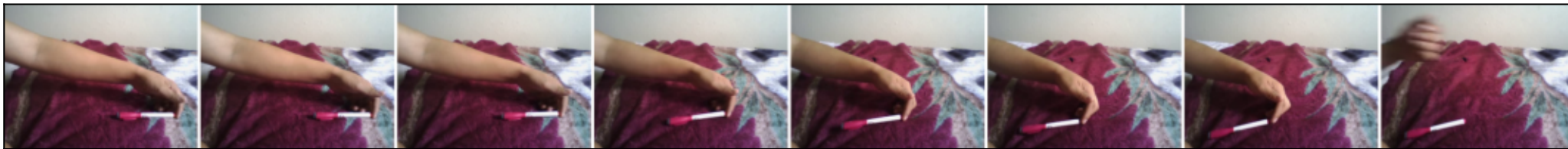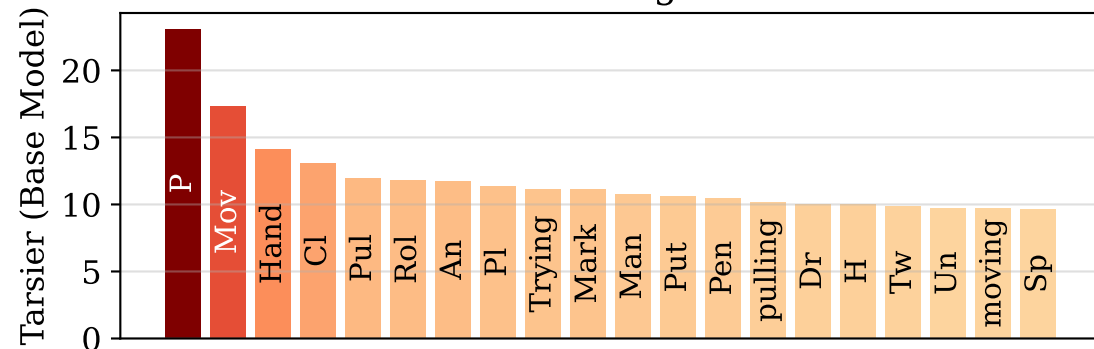Input video with caption: Pushing [something] from right to left

Video logits

Text logits