Method	Average Inference Time per Sentence
MLD	134 ms
MotionDiffuse	6327 ms
MDM	10416 ms
T2M-GPT	239 ms
MoMask	73 ms
Ours	181 ms

Table 1: Computational overhead of different methods.

	FID ↓	Top1 ↑
Cross attention of token and text	$0.034^{\pm.001}$	$0.521^{\pm.003}$
Computing the temporal and spatial attention in parallel	$0.035^{\pm.001}$	$0.522^{\pm.003}$
Ours	$0.033^{\pm.001}$	$0.529^{\pm.003}$

Table 2: Ablation study on HumanML3D dataset.

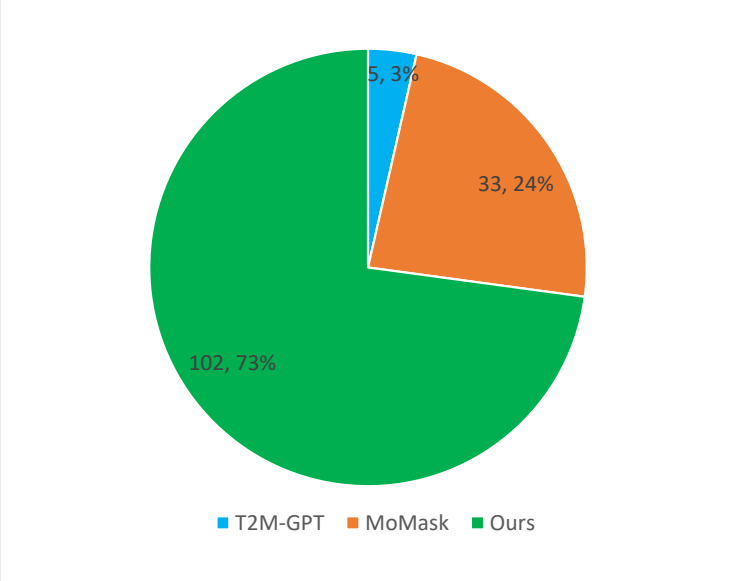

Figure 1: User study.