HW3 – Train virtual humans with PPO

- HW3 asks you to 1) modify the walker's reward settings and train the walker using PPO, and 2) compare and discuss the walker's behavior under different reward settings
- Due: Next class meeting
- Upload ppt to Teams

Reward given every step

void FixedUpdate() {

Add Reward(Match Speed Reward * Look At Target Reward);

$$\left(1-\frac{\Delta v^2}{s}\right)^2$$

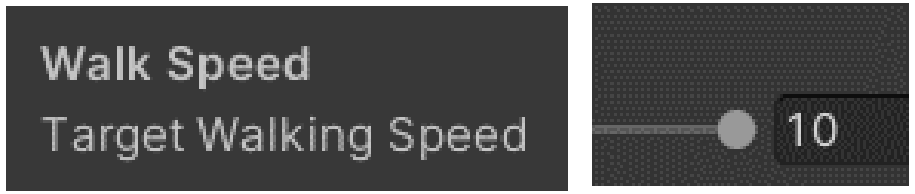
$$0.5 \times (cube_z \cdot head_z + 1)$$

Δv=clamp$\left(\left|v_{goal} - v_{actual}\right|, \text{min}=0, \text{max}=s\right)$

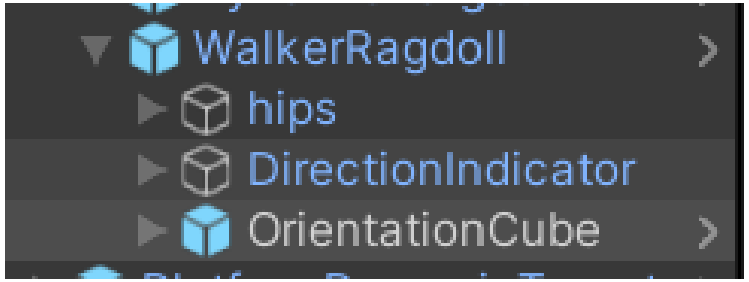
$v_{goal} = cube_z * s$

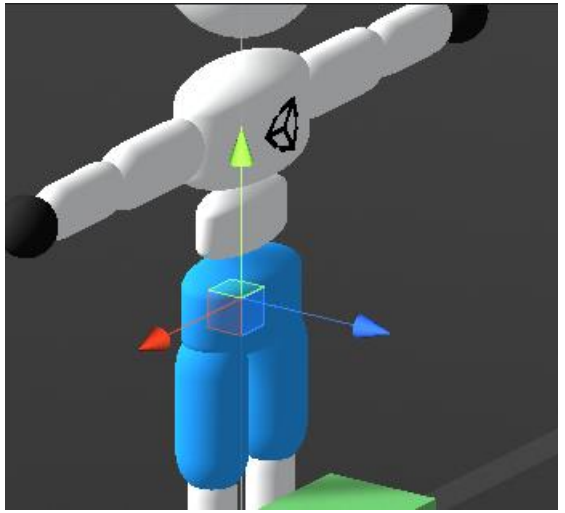
v_{actual} = GetAvgVelocity

s



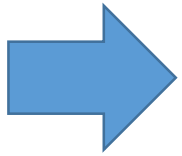
}



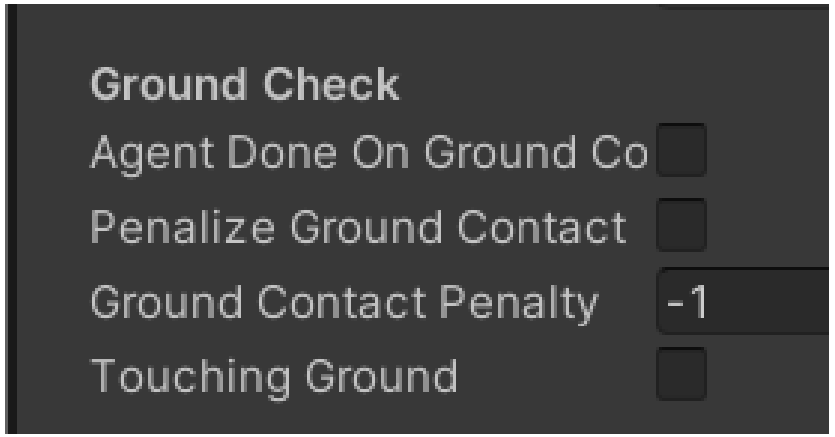


Sparse reward

```
TouchedTarget()
{
    AddReward(1)
}
```



This function is never called?




```
OnCollisionEnter(Collision col)
{
    agent.SetReward(groundContactPenalty);
}
```

Results from original reward settings (HW2)

100K steps

Walker always falls.

500K steps

Walker still can not stand.

1M steps

Walker can stand for a short time.

5M steps

Walker starts to walk without falling for a while with uncertain direction.

6.5M steps

Walker is able to walk better straight.

1105458 Supansa 張淑金

Your reward and PPO training results

Discussion

- Reward settings vs resulted behavior

$$\max J(\pi) = E_{\tau \sim \pi}\left(R(\tau)\right)$$