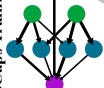


Clotho Test Events not in AudioCaps



AudioCaps Trained

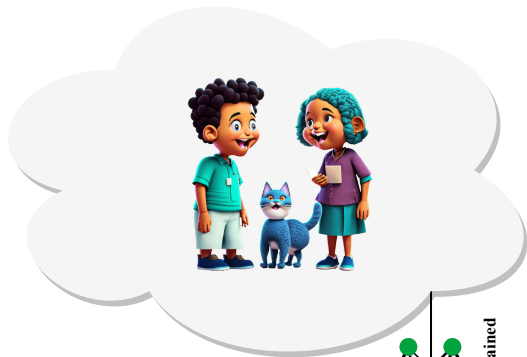


Ground Truth : a jackhammer is being used at an outdoor site while men talk in the distance and a vehicle passes by.

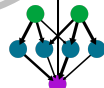
SOTA[6] : a vehicle passes by and passes by.

RECAP (ours) : Someone is hammering and people are talking while a tractor trailer rig drives by.

Compositional Audio



AudioCaps Trained



Ground Truth : a child talking then laughing with a man as an animal gurgles.

SOTA[6] : a baby cries and a woman speaks.

RECAP (ours) : a cat is meowing and people are talking.