



RGB Reference
(Not used
in our method)

Acoustic Signal
(input)

3D Human Pose
(output)



Amplitude

