



Hey Director. Let's create a long-form human speech video using the provided text lines and reference images.



Sure! I will direct the talk by orchestrating **camera shot changes**, **vocal delivery** and **speaker's gestures**.

Reference Images



[1] I'm like, I ain't taking this [Audio Instruction: Annoyed, slightly raised pitch]. My foot said, you better try it... [Camera Instruction: Close Shot]

[2] That's why you see me do this [Motion Instruction: Lift right hand]. I'm making a shoe for people over... [Camera Instruction: Medium Close Shot]

[3] And the only medication they got for gout kills your liver [Audio Instruction: Serious tone, slow pace], so they tell you, take it for a day... [Camera Instruction: Medium Shot]

[4] If you see me dressed up [Motion Instruction: Walk to the left], the first thing I do is I kick my shoes off... [Camera Instruction: Full Shot]



Input Text Lines

[1] I'm like, I ain't taking this. My foot said, you better try it...
[2] That's why you see me do this. I'm making a shoe for people over...
[3] And the only medication they got for gout kills your liver, so they tell you, take it for a day ...
[4] If you see me dressed up, the first thing I do is I kick my shoes off...