

Making AI Voices Cry

A 30-sample systematic experiment with Doubao TTS reveals an undocumented truth: AI voices—even cloned ones—don't just speak. They can feel.



The Emotion Matrix: Heartbreak



Voice trembling with sobs
caught in the throat, crying
while speaking, very sad...

+26% Audio Variation:
Adding the expressive
model drastically
increases pitch and
pacing variation.

Casting Matters:
Female voice 'Su Rou' is
devastating with this
prompt. Test your
specific use case.

The Emotion Matrix: Intimacy

Not all voices respond
equally to intimacy.

Male voice 'Yi Nan' excels
at convincing ASMR.

Always audition your
voices for the required
scene.

ASMR whisper, very
soft and gentle,
pitch rising happily
like acting spoiled,
whispering in the
ear...



✨ The Emotion Matrix: Fatigue



Exhausted, lazy voice, yawning while speaking, soft and squishy tone...

The model acts out physical cues. Adding 'yawning' physically alters the pacing, breath, and audio output of the AI.

The Breakthrough: Paint a Scene

~~Sad~~

(Barely works: The model doesn't respond to labels.)

Verb + Physical Cue + Scenario

(Succeeds: You are directing a performance.)



The single biggest lever in AI TTS isn't the model version or the voice clone. It's the scene you write. Give the AI a scene to act out, not a category to select from.