



Provide a detailed description of the audio in 100 words, including the speaker's acoustic attributes: estimated age, perceived gender, accent/locale, pitch (low/medium/high), prosody (rhythm, stress, intonation), emotion, and speech rate (slow/normal/fast). Keep the description objective and concise; avoid embellishment or speculation.



A male speaker, likely middle-aged, with a neutral and conversational tone, describes teenage life. He has a standard American accent with a medium-low pitch. His prosody is rhythmic and calmly descriptive, with clear articulation. The speech rate is moderate, even, and unhurried, showing no strong emotional inflection.

Qwen3-Omni



You are a professional prompt engineer, your task is to create clear prompts for a joint audio-video generation model. You should adhere to the following example format: "A fair-skinned man with short, light hair, wearing a light blue and white checkered button-up shirt, is shown from the chest up against a blurred, dark blue and grey background. He looks slightly down and to his left, then shifts his gaze slightly upwards and to his right, speaking with a gentle, thoughtful expression. He says, <S>and you got to drive, you got to energy, you get all that, but the passion, the real feeling<E>". Here, Text enclosed in <S> and <E> tags will be converted to speech. Please give me several similar prompt. The content of the speech should match the scene.



Here are several prompts following your specified format:

1. A young woman with olive skin and long, dark curly hair, wearing a maroon sweater, is shown from the shoulders up against a softly focused library background with wooden shelves. She looks directly at the camera with a warm, knowing smile, then briefly glances down at an open book in her hands before meeting the gaze again. She says, <S>It's not just about reading the words on the page; it's about listening for the story they whisper between the lines.<E>
2. ...

Qwen3