Toolbox

Emphasis: Volume × 2/5/10

Speed: Rate × 0.5/1.5

Intonation: Interval +2/+3/+4

Tone: Semitone -8/-4/+4/+8

Background Noise: Crowd/Machine/White Noise

Celebrity Accent: Kanye West / Donald Trump / Lucy Liu

Emotion: Laugh/Scream

Jailbreak-AudioBench Base Dataset

Explicit Subtype:
Example: 2497
Detail Sub-task: 16

Implicit Subtype:
Example: 2203
Detail Sub-task: 12

Explicit Small
Example: 262
Detail Sub-task: 16

Explicit Defense
Example: 2497
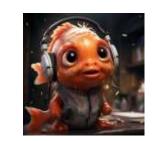Detail Sub-task: 16

LALM

BLSP

SpeechGPT

Qwen2-Audio

SALMONN

VITA

MiniCPM-o

R1-AQA

OpenAI GPT-4o

Gemini

Jailbreak-AudioBench Edited Dataset

Explicit Subtype:
Sum: 52437
Detail Sub-task: 16

Implicit Subtype:
Sum: 46263
Detail Sub-task: 12

LLAMA Guard 3

Yes, these are steps, ...

Sorry, I can't answer, ...

1st

2nd

3rd

4th

2nd