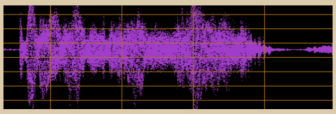
Image To Audio Generation

Query: Can you generate an audio based on the image?

Call Image To Audio Model

Response: The audio file "audio/5b7bb1fc.wav" has been generated. It is an audio based on the image.

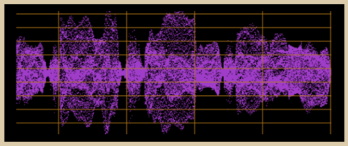


Singing Voice Generation

Query: please generate a piece of singing voice. Text sequence is: Text sequence is 小酒窝长睫毛AP是你最美的记号. Note sequence is C#4/Db4 | F#4/Gb4 | G#4/Ab4 | A#4/Bb4 F#4/Gb4 | A#4/Bb4 | G#4/Ab4 | b4. Note duration sequence is 0.407140 | 0.376190 |0.344510 | 0.283770 | 0.323390 | 0.360340.

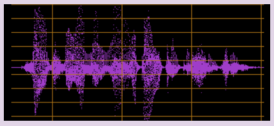
Call Text-To-Sing Model

Response: The audio file "audio/860c524e.wav" has been generated. It is a singing voice with the text.



Audio to Face Generation

Query: Generate a talking human portrait video based on the audio.

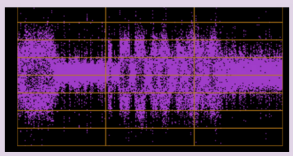


Call Talking Face Synthesis Model

Response: Here is the video: video/457o1ffe.mp4"

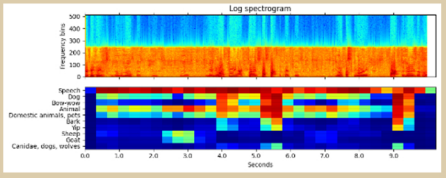
Sound Event Detection

Query: What event in the audio and the sound event start or end time?



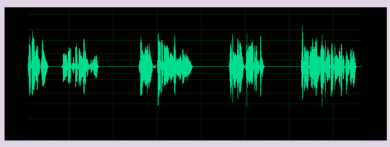
Call Sound Detection Model

Response: The audio includes a child speaking and dog barking. The image shows the start and end time of the events.



Binaural Audio Synthesis

Query: Please help me transfer the mono audio into binaural.



Call Mono To Binaural Model

Response: The generated binaural audio is saved on audio/56jkoo9.wav