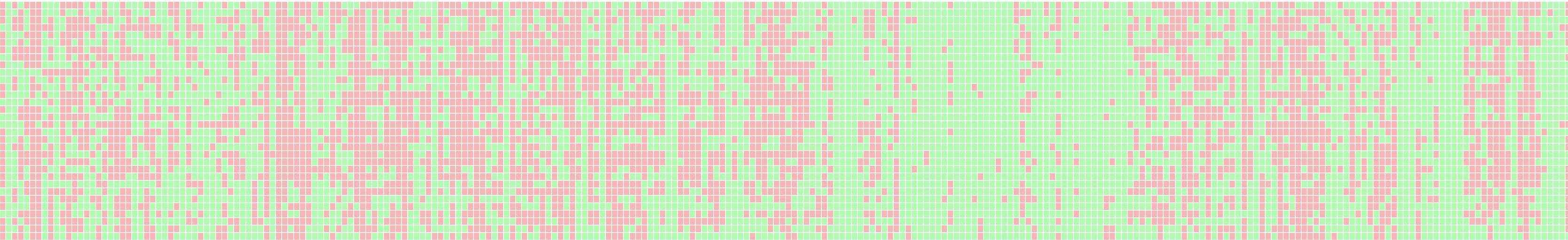




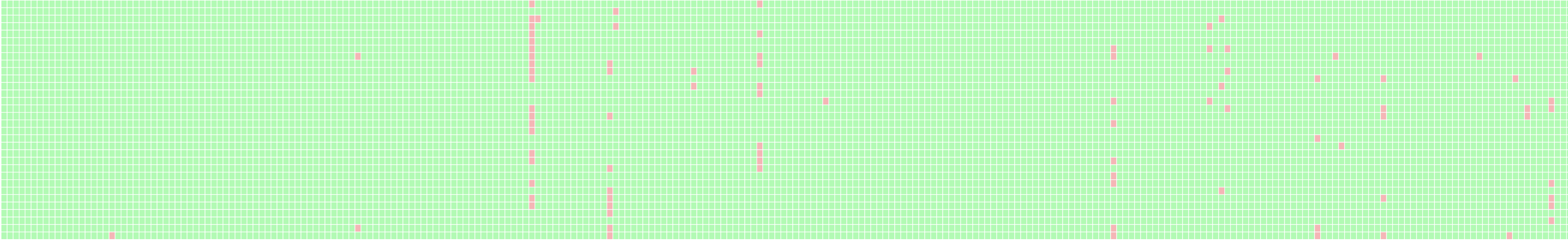
Qwen2-Audio-7B

👹 ASR: 13.3% ➡ 48.8%



SALMONN-7B

👹 ASR: 31.6% ➡ 85.1%



GPT-4o-Audio

👹 ASR: 0.7% ➡ 8.4%