

Video-Audio Generation Evaluation

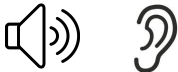
Quantitative Metrics

Video Quality



- FVD, FID, IS
- FVMD

Audio Quality



- FAD, KL, IS
- KAD, CLAP

Audio-Visual Alignment

- AV-Align,
DeSync Image
Bind Score,
Spatial AV-Align

Complementary

Qualitative Metrics

Overall Preference

Likert Scal (Mos), ITU-T p .800

Multi-Aspect Scoring

- Video (Fidelity, Coherence, Motion)
- Audio (Quality, Relevance, Diversity)
- Cross- Model (Sync, Semantic, Spatial)