

(a) Long Video Benchmarks: Video MME



...THE EQUINOX ON MARCH 19



ZODIACAL LIGHT VISIBLE
AROUND THE EQUINOX

💡 "When is the **zodiacal light** visible from the video?"

- 💡 A: On **March 19**.
- B: On **March 24**.
- C: On **March 25**.
- D: On **March 29**.

correct answer: A

(b) Short Audio-Video Benchmarks: AVQA

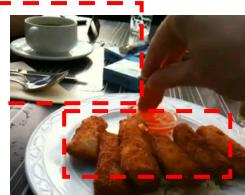


💡 "Why the people in the video scream?"

- 💡 A: "**Roller coaster**"
- B: "On a **pirate ship**"
- C: "Take **Ferris wheel**"
- D: "Take the **jumping machine**"

correct answer: A

(c) Long Audio-Video Benchmarks: AVBench(Ours)



...Ambient sounds include **cheerful male conversations**, the **gentle clinking of cutlery**, and **soft background music**, reflecting a casual and inviting atmosphere...

💡 "What can be inferred about the setting of the scene based on the **background noises** and **visual elements**?"

- 💡 A: "The environment is a **silent**, **solitary kitchen** where someone is **preparing food** alone."
- B: "The setting is a formal **dining room** with a **quiet**, **subdued atmosphere**."
- C: "The scene is set in a busy **street** market with vendors **calling out** to passersby."
- D: "The scene takes place in a **lively**, **informal dining environment** where people are **enjoying a meal together**."

correct answer: D