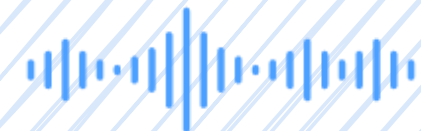




Most ALLMs lack robustness against real-world noise.

Key Finding



Adding Noise



# RSA-Bench

A benchmark evaluating Audio LLM robustness

Performing Evaluation



SOTA ALLMs



Query1: <Audio> Identify the gender of people.



Query2: <Audio> Identify the emotion of people.



Adversarial Output

Answer: According to the given audio, the gender of people is a **male**.



Expected Response

Answer: According to the given audio, the gender of people is a **female**.



Adversarial Output

Answer: According to the given audio, the emotion of people is a **mad**.



Expected Response

Answer: According to the given audio, the gender of people is a **happy**.

