

Question Type: Audio-Visual Temporal

Question: Which ukulele makes the sound first?

Answer: left [LAVisH] simultaneously [Ours] simultaneously [Ground Truth]



Question Type: Audio-Visual Counting

Question: How many types of musical instruments sound in the video?

Answer: five [DG-SCT] six [Ours] six [Ground Truth]