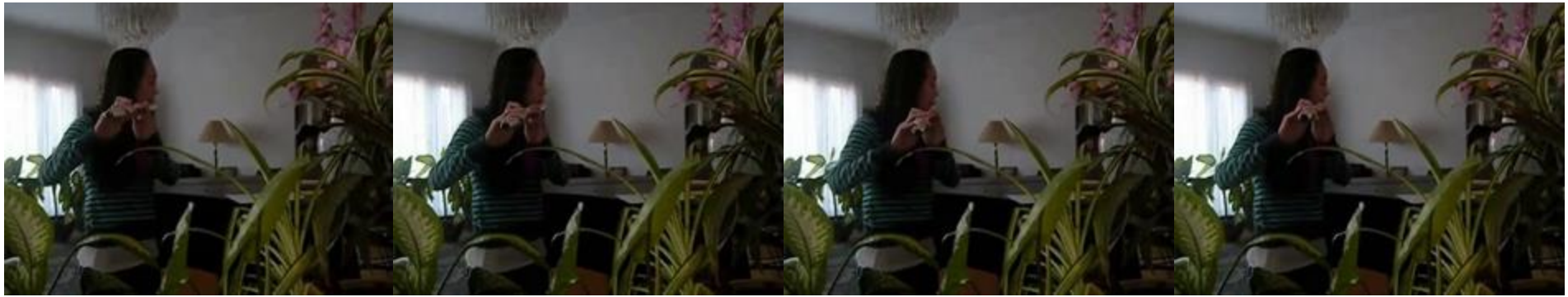
Case 1: Synthetic Video

Question: What is the second instrument that comes in?

Answer: Drum

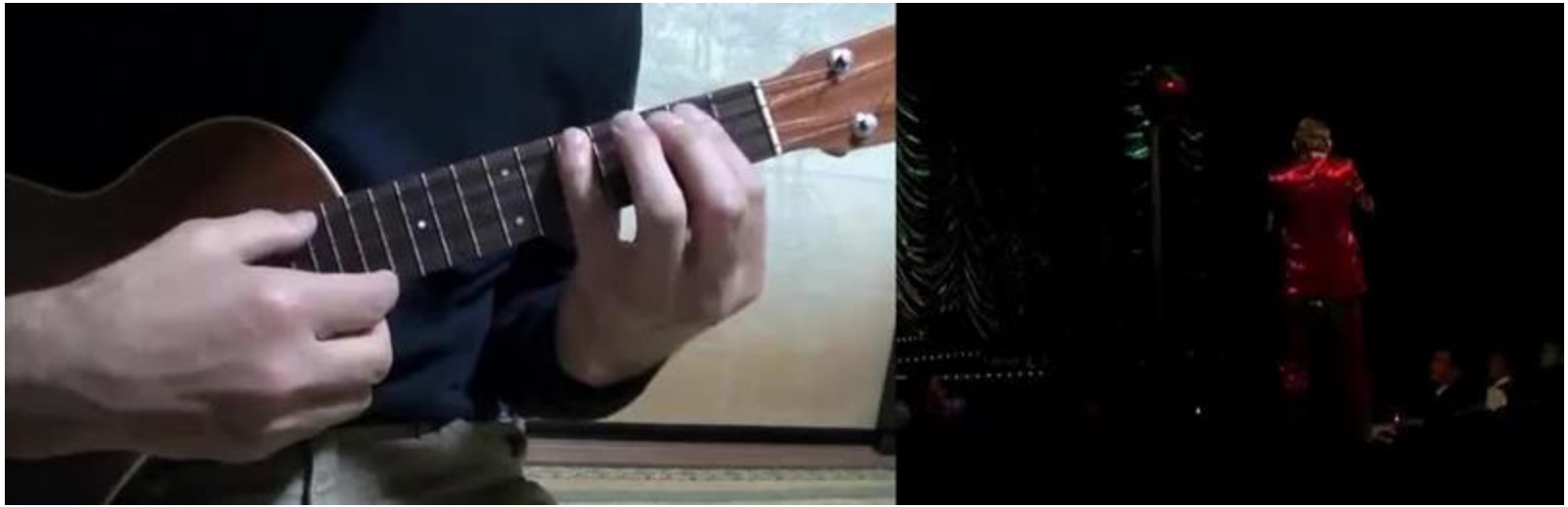
Case 2: Complex Scenes



Question: Where is the performance?

Answer: Indoor

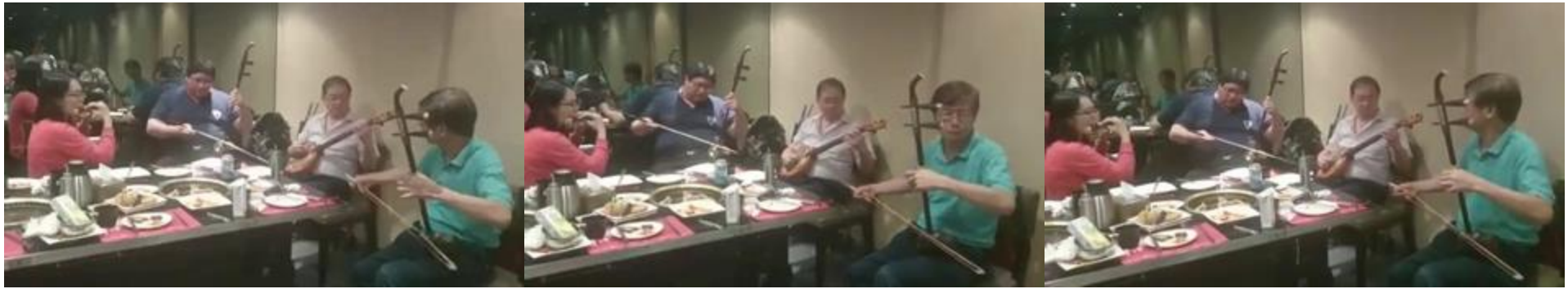
Case 3: Cross-Modality Information



Question: Is the instrument on the right louder than the instrument on the left?

Answer: Yes.

Case 4: Low Resolution



Question: What is the left instrument of the last sounding instrument?

Answer: Erhu.