# Supervised Audio Classification



Audio Encoder → Projection Layers → Class Prob. Vector (1,C)
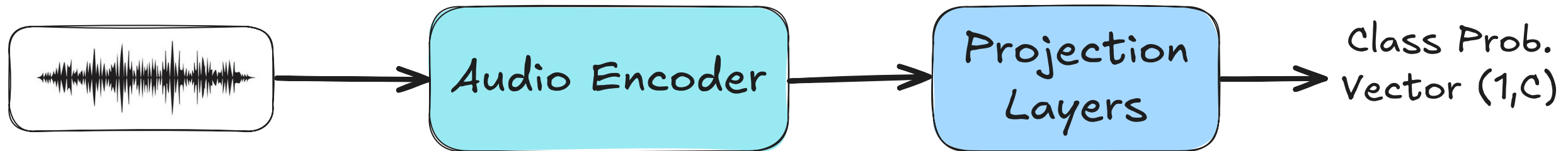
# Zero-shot Audio Classification



Audio Encoder → E_a Embedding

Match

{C1, ... , CM} Classes → {X_t1, ... , X_tM} Prompt Texts → Text Encoder → {E_t1, ... , E_tM} Embeddings