Humans perceive the world using multi-modal sensory inputs such as vision, audition, and touch. In this work, we investigate the cross-modal connection betwe