

Multimodal and Dynamic Sensory Contributions to Adjective Meaning

Do offline sensory norms predict how word meaning is represented in neural dynamics during language comprehension? We address this question by linking sensory norms with time-resolved neural decoding for Chinese adjectives. Although modality-specific norms characterize how concepts relate to different sensory modalities (e.g., vision dominates multisensory processing, Colavita, 1974; Spence et al., 2010; Lynott & Connell, 2009; audition operates as an independent modality with weaker cross-modal coupling, Louwerse & Connell, 2011; smell and taste form a tight integration of chemical senses, Small & Prescott, 2005), it remains unclear whether such self-reported sensory structures are reflected in real-time neural representations. Through two experiments, we show that adjective semantics is multimodal, and that time-resolved neural decoding revealed an early visual-auditory dissociation (~140ms), a later auditory-tactile dissociation (~700ms), and visual-tactile representations showed sustained overlap.

In Experiment 1, 44 native Chinese speakers (with a bachelor's degree or above) rated 65 disyllabic Chinese adjectives (e.g., 光洁 '*glossy*', and 响亮 '*sonorous*') on five sensory dimensions (visual, auditory, tactile, gustatory, and olfactory) using a 0–5 Likert scale. K-means cluster analysis based on mean modality ratings revealed four sensory groups rather than strictly unimodal categories: (a) gustatory–olfactory adjectives ($M_s = 3.48$ and 4.35), (b) visual adjectives ($M = 4.23$), (c) auditory adjectives ($M = 4.04$), and (d) tactile–visual adjectives, characterized by strong tactile ($M = 4.00$) and moderate visual ratings ($M = 2.52$). The results indicate that adjective meanings are best characterized as multisensory representations with different modality prominence.

In Experiment 2, 32 unique participants performed a modality verification task while 64-channel EEG was recorded. Adjectives from the visually, auditorily, and tactually dominant groups identified in Experiment 1 were used. Time-resolved linear discriminant analysis to delta-band oscillatory activity (0.5–4 Hz) (Ding et al., 2016) revealed above-chance discrimination between sensory groups, but with distinct temporal profiles. Visual-auditory distinctions emerged at early time windows, reaching significance around 140ms after stimulus onset (100–210ms), whereas auditory-tactile distinctions emerged at later stages (670–760ms). In contrast, visually and tactually dominant adjectives were not reliably distinguished at any time point.

Together, these findings showed that norm-based sensory structure partially predicts neural representations of word meaning, with modality separability unfolding over distinct time windows. Early visual-auditory distinctions likely reflect their relative independence and frequent use in human experience, whereas later auditory-tactile separation may index more integrative or evaluative processing. Persistent visual-tactile overlap likely to reflect their close coupling in everyday perception. Overall, the results support embodied accounts of lexical semantics, in which sensory-modal contributions to word meaning are graded, interactive, and dynamically accessed based on their roles in human experience in the world. By bridging offline sensory norms with time-resolved neural decoding, the present study showed how multimodal structure in the semantics of adjectives could be dynamically accessed during online language comprehension. Future studies may further test the generality of these findings by examining the spatial organization of sensory semantics and extending the approach to other lexical categories.