Multisensory learning and integration in a first-person fisherman game

Yile Sun\textsuperscript{1}, David C. Somers\textsuperscript{2}, Barbara Shinn-Cunningham\textsuperscript{3} and Robert Sekuler\textsuperscript{4}

\textsuperscript{1}Department of Psychology, Brandeis University
\textsuperscript{2}Department of Psychology, Boston University
\textsuperscript{3}Center for Computational Neuroscience and Neural Technology, Boston University
\textsuperscript{4}Volen Center for Complex Systems, Brandeis University

Integration of information from multiple modalities shapes human cognition, but the limiting factors and mechanisms for such integration are poorly understood. We have examined multisensory integration in a first-person video game in which players encountered rapidly swimming fish belonging to two different species, and made speeded categorization of each fish’s species. In three experiments, fish from the two species were visually identical, but differed in rate at which they oscillated in size. Each fish could be accompanied by a broadband sound that was amplitude modulated at either (i) the same frequency (and in synchrony with) the fish’s size oscillation, or (ii) the frequency that would have matched the visual oscillation of a fish of the other species. This made a fish’s temporal auditory and visual (AV) modulations either Congruent or Incongruent. To increase chances that the AV modulations might be perceptually integrated, we chose frequencies (6 and 8 Hz) similar to speech’s auditory and visual modulations: syllable rate (auditory modulation) and the rate at which the mouth opens and closes (visual modulation). Throughout, players were instructed to categorize fish solely on the visual dimension. Two experiments included control conditions in which each fish was accompanied by no sound or by a broadband sound not temporally modulated. Congruent AV modulation produced faster and more accurate responses than did Incongruent AV modulation. Moreover, control conditions’ responses were indistinguishable from ones with Incongruent AV modulations. Our results demonstrate that temporally-correlated congruent information from multiple modalities can facilitate perceptual-based categorizations.