The good-enough listener: A visual world paradigm reveals the interaction between prediction and bottom-up input

Jiaxuan Li[1], Kayla Keyue Chen[2], Anne Wang[2], Yuhan Shen[2], Yijia Luo[2], Richard Futrell[1], Wing-Yee Chow[2]

[1]University of California, Irvine, [2]University College London

Introduction: Efficient language comprehension involves anticipating upcoming input, but unexpected input can be costly to process. Studies show that this processing cost is reduced when the unexpected word is phonologically related to the predicted word, compared to when it is unrelated. The mechanisms behind this facilitation remain unclear. According to the Spreading Activation Hypothesis, activation of a predicted word may spread to phonologically related alternatives, facilitating the retrieval of unexpected but related input. In contrast, the Misperception Hypothesis suggests that high phonological overlap between the predicted and unexpected word may lead comprehenders to retrieve the predicted word instead of the actual unexpected input, forming a "good-enough" representation. The present study tested these accounts by examining responses to phonologically related versus unrelated unexpected words in a visual-world paradigm. We found evidence for the Misperception Hypothesis.

Method: Participants (n = 44) listened to Chinese sentences while viewing a display of six Chinese words. The sentences strongly predicted a specific word ("book" [书 shu1]) but continued with an unexpected, anomalous word that was either phonologically related to the predicted word through rhyme overlap ("pig" [猪 zhu1]; Experimental condition) or unrelated ("cotton" [棉 mian2]; Control condition). The visual display contained the three critical words and three unrelated distractors. The target word in the Experimental condition was used as the Control in another sentence item, ensuring that the same set of words were used across both conditions. The Spreading Activation Hypothesis predicts facilitated recognition of phonologically related words and thus more fixations on the target in the Experimental condition than in the Control condition. In contrast, according to the Misperception Hypothesis, listeners are expected to persist with their top-down prediction until sufficient evidence overrides it, so it predicts fewer fixations on the target in the Experimental condition than in the Control condition.

Results: More fixation on the predicted word prior to target word onset indicated prediction based on prior contextual information. Following target word onset, participants shifted their gaze away from the predicted word towards the target. Critically, this shift was slower in the Experimental condition than in the Control condition. A cluster-based permutation test was conducted in the post-target-onset (0-1500 ms) time window, using the log ratio of fixations on the target versus the predicted word as the dependent variable. The analysis revealed a significant main effect of condition in the 300-1500 ms interval (cluster mass = 406, $p < .001$), with lower log ratios observed in the Experimental condition than in the Control condition. No significant differences in log ratios between conditions were found prior to target onset.

Conclusion: Using a visual world paradigm, we found that listeners were less likely to fixate on the target (and more likely to fixate on the predicted word) when the presented target was phonologically related to the predicted word than when it was unrelated. This finding suggests that strong top-down predictions can override bottom-up input, leading to inaccurate representations when the input is confusable with a more predictable alternative.