

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 Department of Language Science, University of California Irvine

2 Division of Psychology and Language Sciences, 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. The Spreading Activation Hypothesis posits that activation of a predicted word spreads to phonologically related alternatives, facilitating their recognition. In contrast, the Misperception Hypothesis suggests that high phonological overlap may cause listeners to misidentify the input as the predicted word, forming a “good-enough” representation. This study tested these accounts using a visual-world paradigm. **Method:** Forty-four participants listened to Chinese sentences that strongly predicted a specific word (e.g., “book” [书 shu1]) but instead continued with an unexpected, anomalous word that was either phonologically related through rhyme overlap (e.g., “pig” [猪 zhu1]; Experimental) or unrelated (e.g., “cotton” [棉 mian2]; Control). While listening to the sentences, participants simultaneously viewed a visual display containing the three critical words and three unrelated distractors. The Spreading Activation Hypothesis predicts *more* fixations on the target in the Experimental condition, whereas the Misperception Hypothesis predicts *fewer*. **Results:** Before target onset, participants preferentially fixated the predicted word, indicating anticipatory processing. After target onset, gaze shifted towards the target from the predicted word. A cluster-based permutation test on the log ratio of fixations (target vs. predicted word) revealed a significant effect of condition in the 300-1500 ms window (cluster mass = 406,  $p < .001$ ), with *lower* ratios in the Experimental condition. **Conclusion:** These findings indicate that phonological similarity between the predicted and actual input interferes with the recognition of the actual input, supporting the Misperception Hypothesis. Strong top-down predictions can override bottom-up input, leading to inaccurate representations when the input resembles a more predictable alternative.

**Corresponding author:** Jiaxuan Li, Kayla Keyue Chen

**Email Address:** [jiaxul19@uci.edu](mailto:jiaxul19@uci.edu), [keyue.chen.19@ucl.ac.uk](mailto:keyue.chen.19@ucl.ac.uk)

Either **ORAL** or **POSTER** presentation