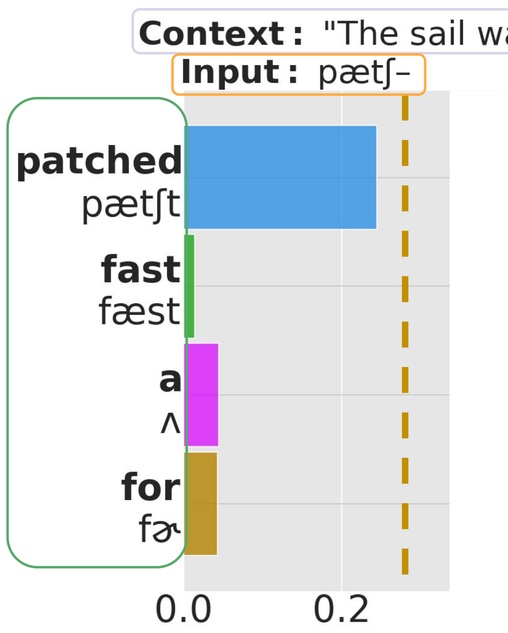
The neural dynamics of word recognition and integration

Listeners **recognize** and **integrate** spoken words by combining expectations about upcoming content with acoustic input.

We model this process of **recognition** and the downstream neural correlates of **integration** in EEG data recorded as subjects listened to naturalistic English speech [1], and ask:

- Is integration tightly yoked to the timing of word recognition, or are they independent?
- Do listeners integrate words differently depending on how easy they are to recognize?

WORD RECOGNITION MODEL



We estimate a listener's beliefs over the speaker's intended word as a function of *context* and *incremental input*.

We say a word is *recognized* just when its probability exceeds a threshold parameter ***T***.

$$P(\text{Word} \mid \text{Context}, \text{Input}) \propto P(\text{Word} \mid \text{Context})\, P(\text{Input} \mid \text{Word}) \geq T$$

How likely am I to hear this **Word** in this **Context**?
Ask a language model.
(GPT-Neo; Black et al. 2021)

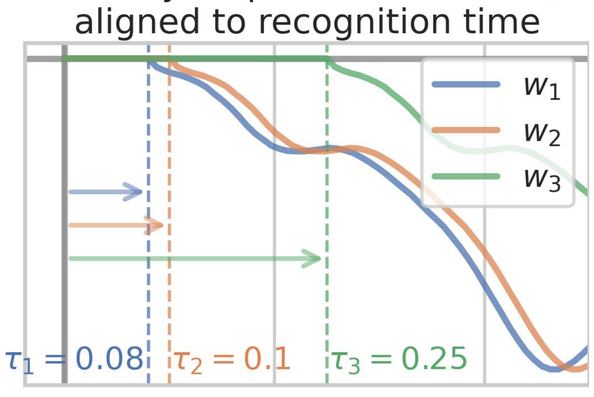
How likely is a **Word** to be realized as **Input**?
Estimate from human confusion data.
(Cutler, Weber, Smits, & Cooper 2003)

NEURAL LINKING MODELS

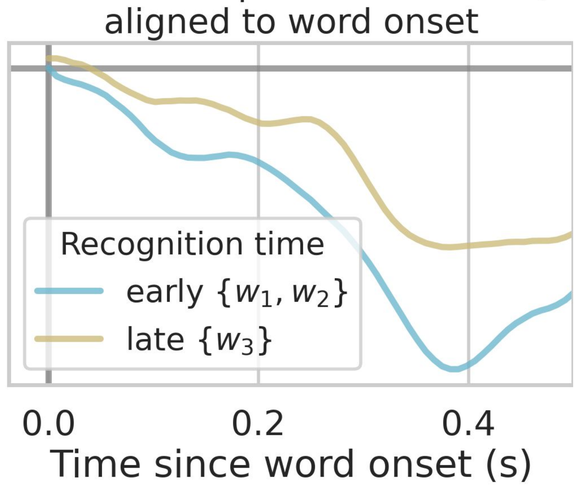
We design variants of the temporal receptive field model (TRF; [2]) which define how the neural response to a word's *surprisal* depends on the word's recognition time.

We estimate these TRF parameters (together with the recognition model) to predict EEG data [1].

Shift :
Unitary surprisal modulation, aligned to recognition time



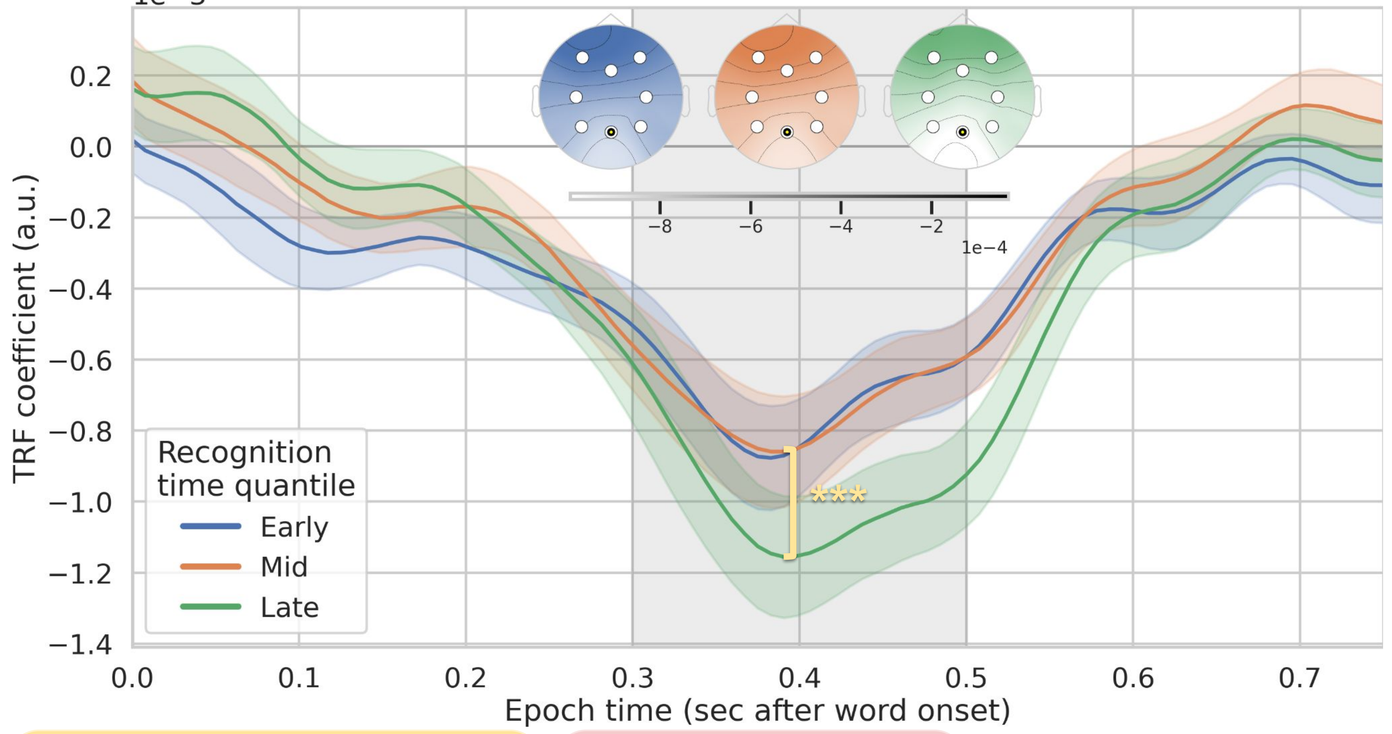
Variable :
Variable surprisal modulation, aligned to word onset



RESULTS

The **Variable** model better predicts held-out EEG data than a baseline not incorporating recognition dynamics ([1]; p < 4e-6), while the **Shift** model does not (p > 0.5). We next investigate the internals of the optimal **Variable** model:

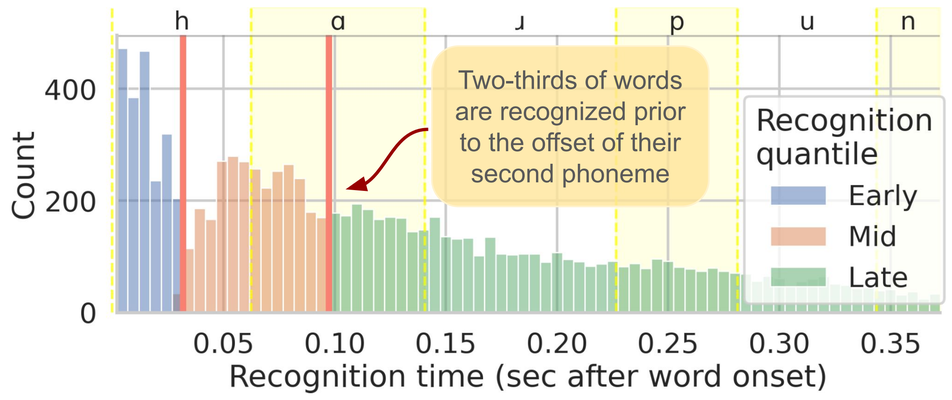
Estimated EEG modulation due to word surprisal, by recognition time

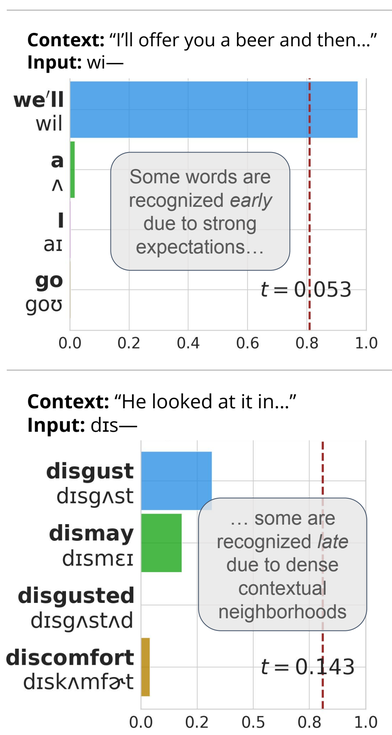


Surprisal modulation significantly **amplified** for **late**-recognized words
within-subject t = -5.14, p < 0.0001

No latency difference to peak negative modulation
within-subject t = 1.391, p > 0.18

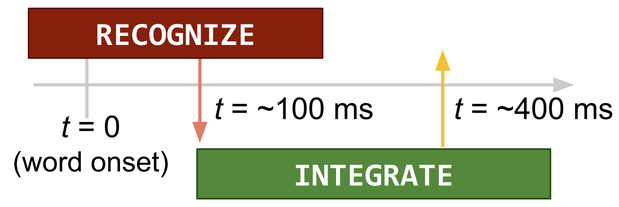
Distribution of estimated word recognition times

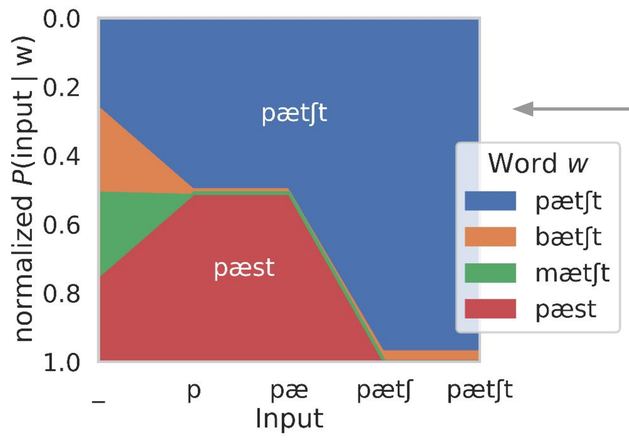




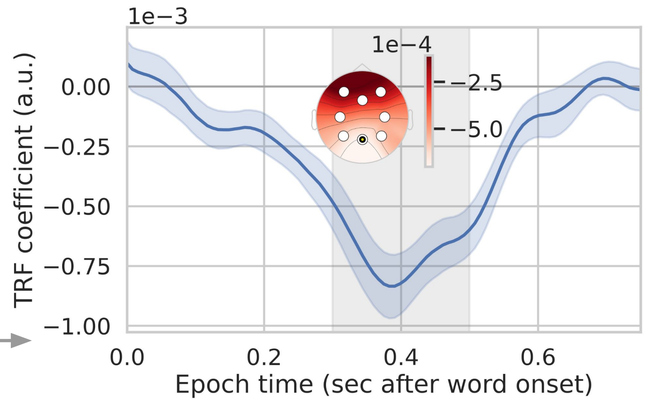
TAKEAWAYS

Word integration is time-locked to word *onset*, not word *recognition time*. Integration shows amplified neural dynamics for *late-recognized* words.





The recognition model likelihood defines how likely any word (colors) is to be realized as noisy, incremental phonetic input (horizontal axis).



N400: EEG event-related potential peaking 400 ms after a word's onset, reflecting the difficulty of lexico-semantic integration; amplitude well predicted by the word's surprisal

[1] Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences.*
[2] Lalor, E. C., Power, A. J., Reilly, R. B., & Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology*.

Take a picture to download the **full paper**

Jon Gauthier and Roger Levy