The Cognitive Throttle of Language: Exploring the Limits of Information Processing

Chia-Wen Lo[1], Lena Henke[1], & Lars Meyer[1,2]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE
[2]Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, DE

lo@cbs.mpg.de

MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES

LANGUAGE CYCLES MAX PLANCK RESEARCH GROUP

Introduction

Temporal constraints

- Auditory short-term memory is limited to 2-3 seconds [1].
- For language, a proposed window of 2.4 seconds, including ~6 words when assuming a rate of 150 words per minute [2-3].
- Timing constraint may come from limited duration of underlying electrophysiological windows: cycles of low-frequency neural activity serve the formation of multi-word chunks.
  - Phase angles of oscillatory activity in the delta band (< 4 Hz) predict the offsets of multi-word chunks [4], in particular when chunks lasts for 2.7 seconds [5].
- Yet, time window is confounded with amount of information

Uniform Information Density Hypothesis

- Information density: the amount of information in a unit
- Uniform information density (UID): speakers prefer utterances that convey/distribute information uniformly across speech signals [6-7].
- Some psycholinguistic evidence for UID:
  - Similar information rates (~39 bits/s) for syllables across different languages [8-9].
  - Entropy rate increases with sentence number [10].

Main questions

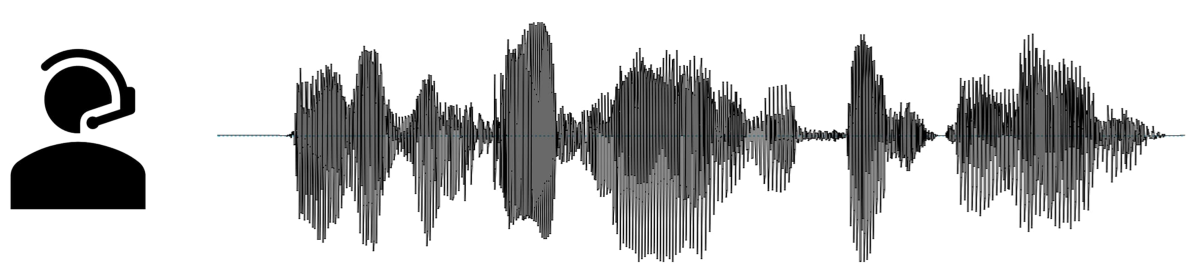
- **According to UID, is there a limit to the amount of information that can determine how we define a chunk in continuous speech?**
- **In particular, does the chunk-related neural activity correlate with chunk boundaries defined by the summation of surprisals?**

Methods

EEG preprocessing

- Analysis of openly available dataset [11]
- 18 native English speaking young adults (19–38 years old)
- Electroencephalography (EEG) recording during naturalistic story listening ("The Old Man and the Sea" by Ernest Hemingway)
- Automated EEG pre-processing (adjusted from HAPPE; [12])

Information in a chunk

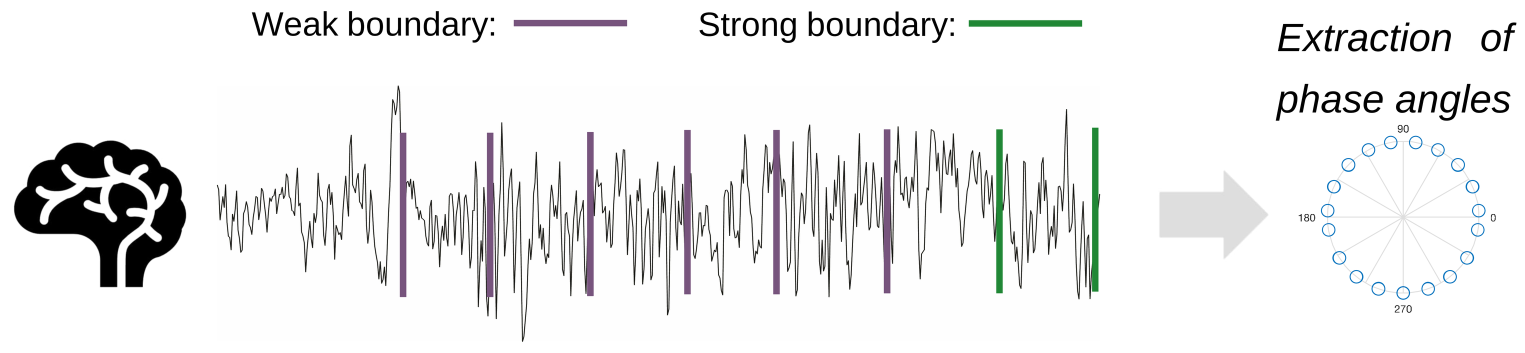


Text:	He	was	an	old	man	who	fished	alone…
GPT2 surprisal:	*8.58*	*3.31*	*6.03*	*5.98*	*1.75*	*2.62*	*14.52*	*8.33…*
Cut-off (5):	*1*	*0*	*1*	*1*	*0*	*0*	*1*	*1…*
Cut-off (10):	*0*	*1*	*0*	*1*	*0*	*0*	*1*	*0…*
…					*…*			*…*
Cut-off (50):	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1…*
SUM:	*1*	*1*	*2*	*3*	*1*	*0*	*6*	*5…*

Median split (SUM>3): to determine the most possible boundaries

EEG analysis

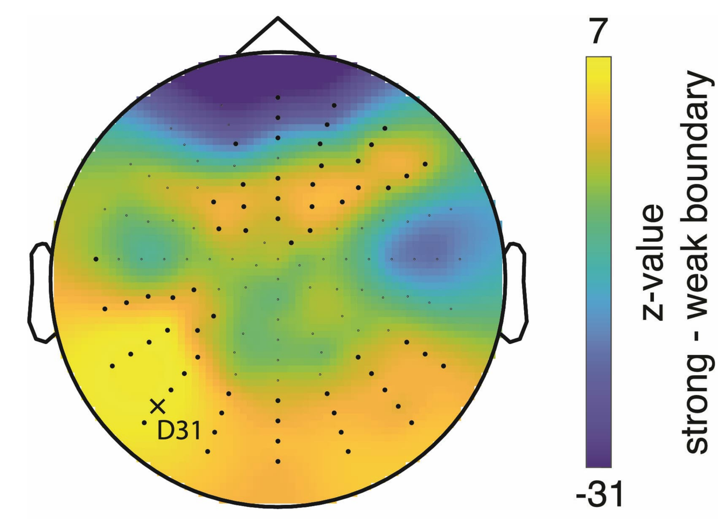
- Extraction of phase angles (< 2 Hz) at word-offsets



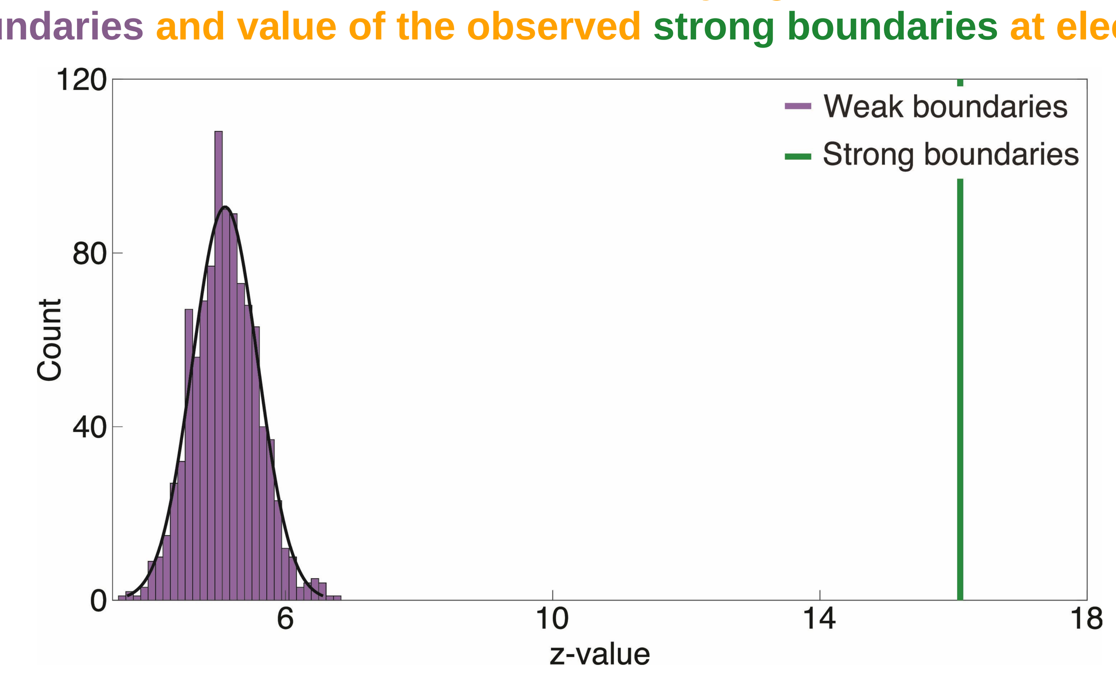
- Comparison of phase clustering at surprisal (strong) boundaries against a surrogate distribution based on word-offsets of weak boundaries

Results

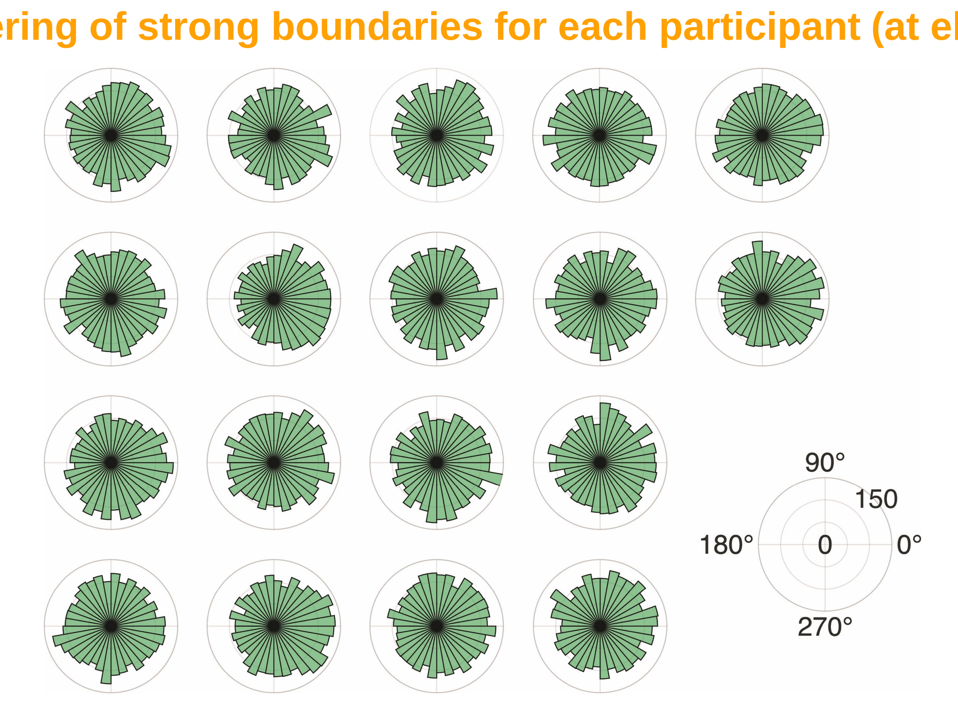
Topography of phase clustering difference between strong and weak boundaries (Median Split: SUM >3)



Histogram of the statistical values from the Rayleigh's test of the boootstraped weak boundaries and value of the observed strong boundaries at electrode D31



Phase clustering of strong boundaries for each participant (at electrode D31)



Discussion

Summary

- Phase clustering of the delta band (< 2 Hertz) is higher for the strong boundaries defined by the surprisal cut-offs, as compared to the weak boundaries.
- This suggests after accumulating enough information (i.e., surprisal values), participants insert a chunk boundary.
- Limitations of information processing indeed play a role in determining multi-word chunks and are reflected by neural processing windows.

Future directions/open questions

- Different ways of quantifying information density in chunk (e.g. local vs. global, [13-15])
- How to tease apart timing constraints and information density in a chunk
- How to determine optimal surprisal value (i.e., which amount of information can be processed best)
- In addition to GPT-models, there are various ways of computing surprisal values (e.g., syntactic surprisal to account for linguistic structure, [16])
- If information is distributed uniformly, chunks should not have a pre-determined beginning or end (e.g., surprisal in chunks may be uniform within a moving window)

References

[1] Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.
[2] Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model*. Cognition, 6(4),* 291–325.
[3] Tauroza, S. & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*, 90-105.
[4] Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex, 27(9),* 4293– 4302.
[5] Henke, L, & Meyer, L. (2021). Endogenous oscillations time-constrain linguistic segmentation: cycling the garden path. *Cerebral Cortex, 31*:4289–4299.
[6] Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.
[7] Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61(1),* 23–62
[8] Coupé C, Oh YM, Dediu D, & Pellegrino F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances, 5(9):*eaaw2594.
[9] Pellegrino F., Coupé C., & Marsico E. (2011). Across-language perspective on speech information rate. *Language, 87*, pp. 539-558.
[10] Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 199–206.
[11] Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, & Lalor EC. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current Biology, 28(5):*803-809.e3.
[12] Gabard-Durnam LJ, Mendez Leal AS, Wilkinson CL, & Levin AR. (2018). The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Front. Neurosci.12*:97.
[13] Meister C., Pimentel T., Haller P., Jäger L, Cotterell R., & Levy R. (2021). Revisiting the Uniform Information Density hypothesis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 963–980, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
[14] Jain A., Singh V., Ranjan S., Rajkumar R., & Agarwal S. (2018). Uniform Information Density effects on syntactic choice in Hindi. In *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing*, pages 38–48. Association for Computational Linguistics.
[15] M Richter, MBI Farré, M Kölbl, Y Kyogoku, JN Philipp, T Yousef, G Heyer, & NP Himmelmann. (2022). Uniform Density in Linguistic Information Derived from Dependency Structures. *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022)* - Volume 1, pages 496-503.
[16] Brennan JR, & Hale JT. (2019). Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE 14(1):* e0207741.

SNL 2023