Time-course of processing for syntactic properties in spoken word recognition

Alexa S. Gonzalez[1,2], McCall E. Sarrett[1], and Joseph C. Toscano[1]
[1]Villanova University, [2]University of Houston

Introduction

- A fundamental issue in spoken language comprehension involves understanding the interaction of linguistic representations across different levels of organization (e.g., phonological, lexical, syntactic, and semantic; Folk, 2017; Trueswell et al., 1994)
- There is debate about when different levels are accessed during spoken word recognition
- Serial processing models predict that comprehension is sequential

Acoustic Cue Encoding → Phonological Processing → Lexical Access → Syntactic/Semantic Procesisng

- Parallel processing models predict simultaneous activation of representations at multiple levels
- Current study isolated neural responses to one type of higher-order information—syntactic class of a word (nouns vs. adjectives)—from low-level acoustic and phonological responses using a component-independent event-related potential (ERP) design
- We predicted that parallel procesing would be revealed by overlap in the time-course across different levels of linguistic organization
  - Previous work has demonstrated that acoustic cue encoding occurs 100-200 ms post-stimulus onset (Toscano et al., 2018)
  - Simultaneous processing of syntactic class information would be observed as differences between nouns and adjectives within 200 ms after the point of disambiguation (POD)

Method

Participants: N=26 listeners; age range: 18-22 years; 42% female; self-reported normal hearing, normal or corrected-to-normal vision, and English as their native language

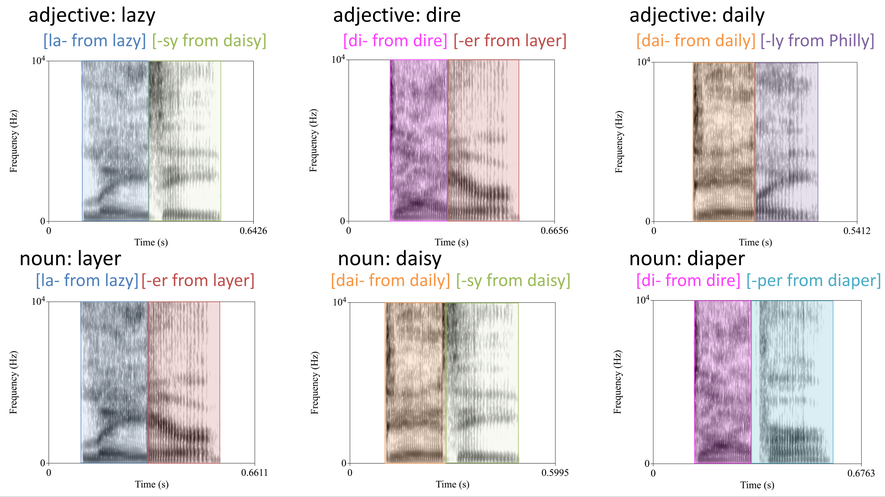
Stimuli
- 20 synthesized disyllabic words (10 nouns and 10 adjectives)
- Noun and adjective lists matched for word frequency, phonological neighborhood density, and biphone probability
- Synthesized using the voice "Samantha" from Apple Text-to-Speech
- Stimuli cross-spliced at syllable boundary (mean POD: 193 ms)

EEG Recording
- 32 electrodes placed at International 10-20 System sites
- Electrode impedances less than 10 kΩ
- Data recorded at a sampling rate of 500 Hz, referenced online to the left mastoid, and re-referenced offline to average mastoids

Procedure
- On each trial, participants heard a spoken word and then saw a visually presented word 500±150 ms later
- Participants indicated whether the auditory and visual words shared the same syntactic class (match vs. mismatch)
- Each combination of 20 auditory and 20 visual items was included, except when the two words were the same, for a total of 380 trials

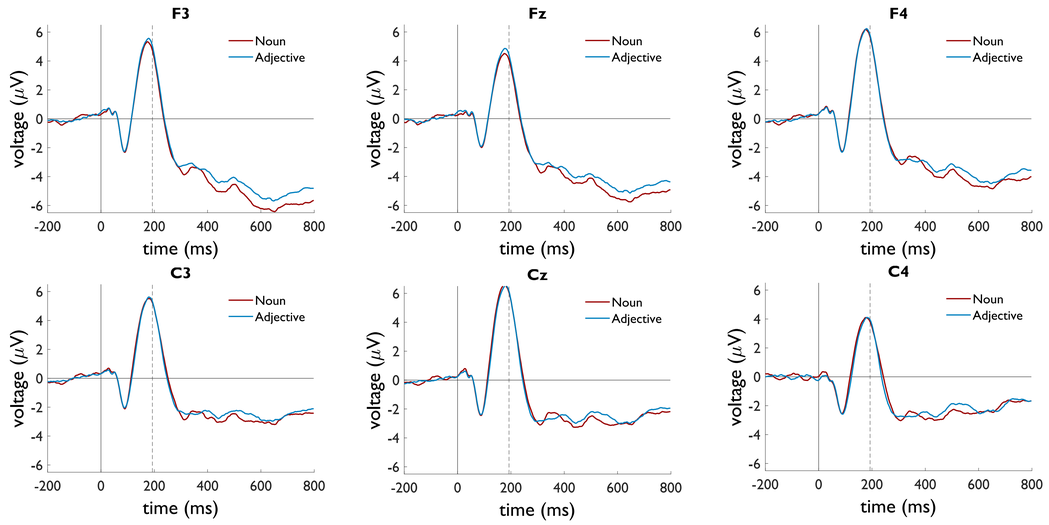
Stimuli

- Examples of stimuli spliced at syllable boundary
- Onset and offset segments matched across the set of nouns and adjectives, canceling out low-level acoustic differences between the two sets of stimuli (see Baart & Samuel, 2015, for a similar approach)

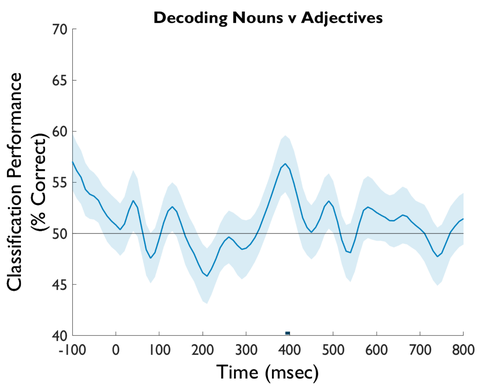


Results

Grand average ERP waveforms at F3, Fz, F4, C3, Cz, and C4; dotted lines represent POD



Decoding Analysis
- Support-vector machine (SVM) trained on two-alternative forced choice classification job (noun vs. adjective)
- Input to SVM was the averaged mean voltage in a 10-ms time window
- Cross validation was performed using a 3 k-fold procedure
- Free parameters of the SVM (*C* and *γ*) optimized at 350 ms post-target word onset
- SVM was run from -100 to 800 ms in 10-ms intervals
- Data smoothed with a 50-ms triangular window



Discussion

- ***Mean Amplitude Analysis:*** Difference between noun and adjective waveforms at 400-450 ms post-target word onset (207-257 ms post-POD) at Fz ($t(25) = -2.14$, $p = 0.04$)
- ***Decoding Analysis:***
  - Statistical significance evaluated with t-tests against numerical chance (Sarrett & Toscano, submitted), using Bootstrapped Difference of Timeseries (bdots) to correct for multiple comparisons (Seedorff et al., 2018); autocorrelation of the test statistic was 0.90, and new alpha level was 0.0074
  - Above chance performance seen at 390 ms post-target word onset (197 ms post-POD)
- ***Conclusions:*** Results provide preliminary support for parallel processing models of spoken word recognition in which higher-level information is accessed while early acoustic analysis is still occurring
- ***Future Directions***
  - Examine processing of semantic class based on cosine similarity between items in a word embedding model
  - Is there evidence of early processing of semantic effects and is the timing similar to timing for syntactic properties of words?

Acknowledgements and References

We would like to thank Olivia Montañez for assisting with pre-processing of the ERP data, and all the members of the WRAP Lab for help with data collection.

This material is based on upon work supported by the National Science Foundation under Grant No. 1945069. This work used the Augie High-Performance Computing cluster, funded by NSF Grant No. 2018933, at Villanova University.

References:

Baart, M., & Samuel, A. G. (2015). Early processing of auditory lexical predictions revealed by ERPs. *Neuroscience Letters, 585*, 98-102.

Folk, E. (2017). *Parallel vs. serial processing in language comprehension*. [Master's thesis, Villanova University].

Sarrett, M.E. & Toscano, J.C. (submitted). Decoding speech sounds from neurophysiological data: Practical considerations and theoretical implications.

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language*, *102*, 55–67.

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language, 184*, 32-42.

Trueswell, J.T., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285-318.