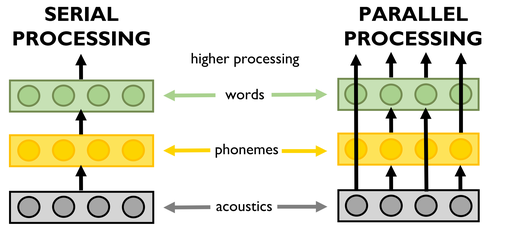
1pSC10

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

Olivia Montañez[1,*], McCall E Sarrett, PhD[1,*], Alexa S Gonzalez, MS[2], & Joseph C Toscano[1]
[1]Villanova University, [2]University of Houston
[omontane,mccall.sarrett]@villanova.edu

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



Predicts that comprehension is **sequential**
Frazier & Fodor (1978)
Frazier & Clifton (1996)

Predicts that multiple levels of processing occur **simultaneously**
Tanenhaus et al. (1995)
Trueswell et al. (1994)

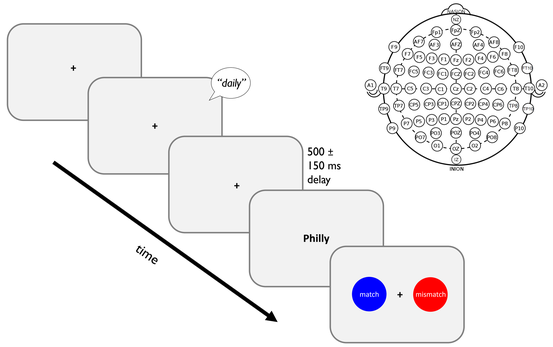
There is debate about when information at different levels is accessed during spoken word recognition

Current study isolated neural responses to one type of higher-order information—syntactic class (nouns vs. adjectives)—from low-level acoustic and phonological responses using a component-independent event-related potential (ERP) design

We predict overlap in the time-course across different levels of organization, supporting parallel processing models

- Previous work has shown that acoustic cue encoding occurs 100-200 ms post-stimulus onset Toscano et al., 2018
- Simultaneous processing of syntactic class information would be observed within 200 ms after the point of disambiguation

Current Study



Method

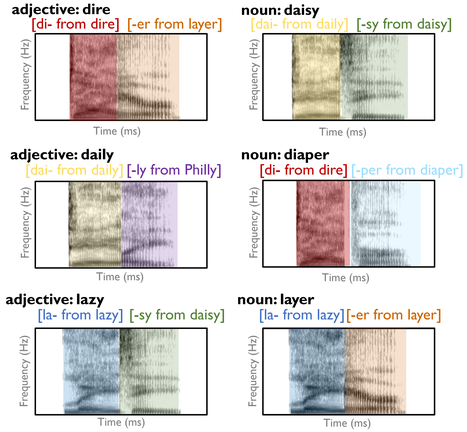
Participants: N=34 listeners; age range: 18-22 years; 16 female, 18 male; self-reported normal hearing, normal or corrected-to-normal vision, and English as their native language

Design

- On each trial, participants heard a spoken word and saw a visually presented word 500 ± 150 ms later, while EEG data were recorded
- Participants indicated whether the auditory and visual words shared the same syntactic class using a button box
- 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

- 20 synthesized disyllabic words (10 nouns and 10 adjectives)
- Lists matched for frequency, phon. 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)
- Onset and offset segments matched across set of nouns and adjectives, canceling out acoustic differences between the two sets of stimuli see Baart & Samuel, 2015, for a similar approach



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

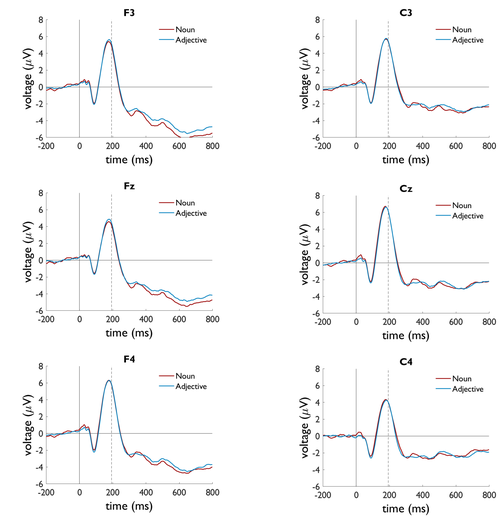
EEG Pre-processing

- MATLAB was used to convert raw EEG data into ERP waveforms: re-referenced to average of left and right mastoids, sequentially high- and low-pass filtered from 0.1 to 30 Hz, nonstereotypic artifacts rejected (ICA corrected for eyeblinks as needed), timelocked to auditory word, epoched from -200 to 800 ms with 200 ms pre-stimulus baseline, averaged ERPs computed

Results

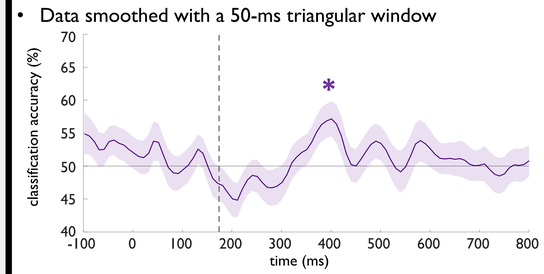
ERP waveforms

- 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 window
- Cross validation was performed using a 3 k-fold procedure
- Free parameters 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



- Statistical significance evaluated with t-tests against numerical chance Sarrett & Toscano, submitted
- BDOTS to correct for multiple comparisons Seedorff et al., 2018
- $\rho = 0.93$, and new $\alpha = 0.0068$

Discussion

Conclusions

- Above chance performance seen at 390 ms post-target word onset (197 ms post-POD)
- Results provide support for parallel processing models of spoken word recognition in which higher-level information is accessed while early acoustic analysis is still occurring
- Syntactic class of words is accessed early and concurrently with acoustic processing

Future Directions

- Examine processing of semantic class based on cosine similarity between items in a word embedding model
- Examine differences in lexical status (words vs. nonwords)
- Is there evidence of early processing of lexical or semantic effects, and is the timing similar to that for syntactic properties of words?

References

Folk, E. (2017). *Parallel vs. Serial Processing in Language Comprehension.* Master's Thesis, Villanova University.

Frazier, L., & Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291-325.

Frazier, L., & Clifton, C. (1996). *Construal.* MIT Press.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. doi:10.1126/science.7777863

Sarrett & Toscano (under review). Decoding speech sounds from neurophysiological data: practical implications and theoretical considerations.

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.

Trueswell, J.C., 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*(3), 285-318.

Acknowledgments

We would like to thank all the members of the WRAP Lab for help with data collection for this experiemnt.

This material is based 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.