Encoding acoustic cues in speech: Electrophysiological measurements

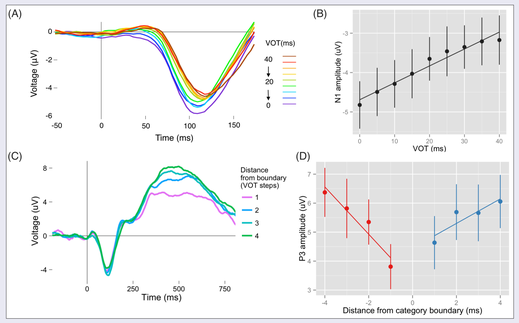
APCAM 2017

Agnes Gao[1] [ygao2@villanova.edu], Olivia Pereira[2] [opereira@villanova.edu], Joe Toscano[1] [joseph.toscano@villanova.edu]

[1] Villanova University, [2] Nemours Alfred I. duPont Hospital for Children, Nemours Biomedical Research

INTRODUCTION

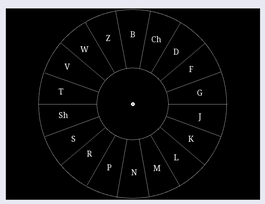
- Early work suggested fine-grained acoustic differences are discarded early in perception[1], but more recent evidence has refuted this claim[2,3]
- However, it is unclear how the perceptual system encodes specific acoustic cues at early stages
- Previous work presented an ERP approach to measure early perceptual responses,[4] examining differences along VOT continua:
  - P300 varies as a function of graded acoustic changes, relative to phoneme categories
  - Auditory N100 varies linearly with changes in VOT, suggesting a measure of early cue encoding independent of phoneme categories



- **Questions:** Does N100 serve as an index of cue encoding for other cues and phonological contrasts? What can this tell us about perceptual organization?

METHOD

- **Exp. 1:** Word-initial minimal pairs spanning 18 consonants (/b,d,g,p,t,k,f,v,s,z,ʃ,m,n,tʃ,dʒ,l,r,w/); **Exp. 2:** Naturally-produced sounds varying along VOT and release burst continua (/b,d,g,p,t,k/)
- Indicated which sound each word began with by clicking corresponding letter(s) on a display

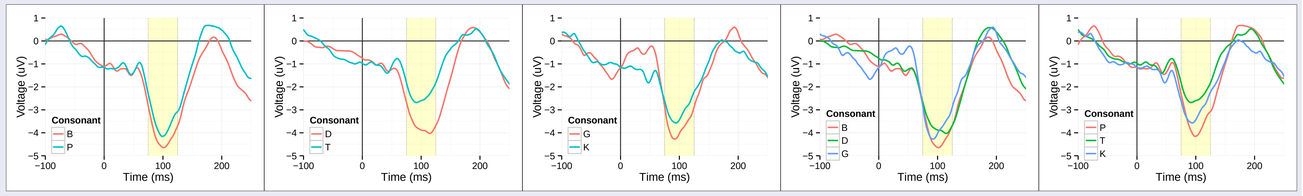


- EEG recorded using 64-ch Brain Products actiCHamp system with electrodes at standard 10-20 sites; N100 measured from average of F3, Fz, and F4
- EOG recorded via electrodes horizontal to the external canthi of each eye (horizontal) and an electrode located above the left eye (vertical); trials containing oculomotor and muscle artifacts rejected via a peak-to-peak threshold detection and visual inspection
- Electrode impedances kept at <10 kΩ; Data referenced to left mastoid online and rereferenced to average mastoids offline; EEG recorded at a sampling rate of 500 Hz; offline bandpass filtered at 0.1 to 30 Hz in ERPLAB (Butterworth filter with 12 dB/octave roll-off)

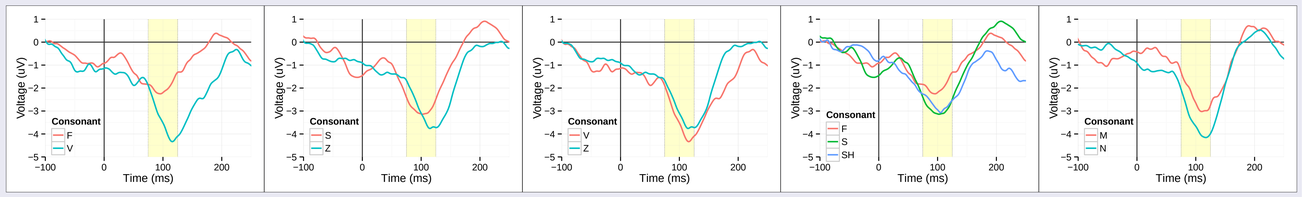
RESULTS—EXPERIMENT 1 (Natural speech)

Grand average ERP waveforms (F3,Fz,F4)

Stop consonants

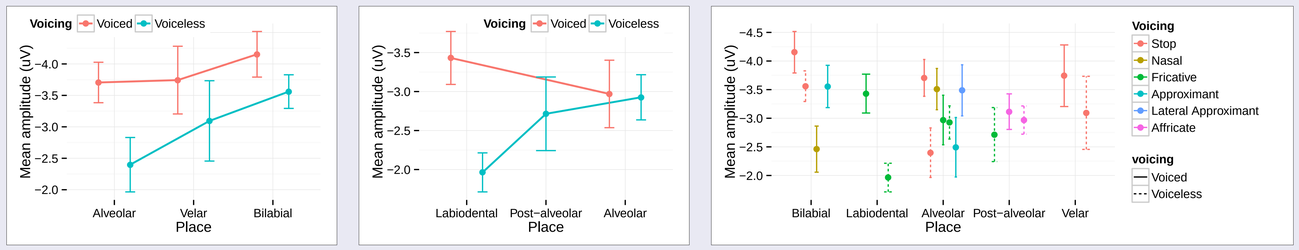


Fricatives



Nasals

Mean N100 amplitude



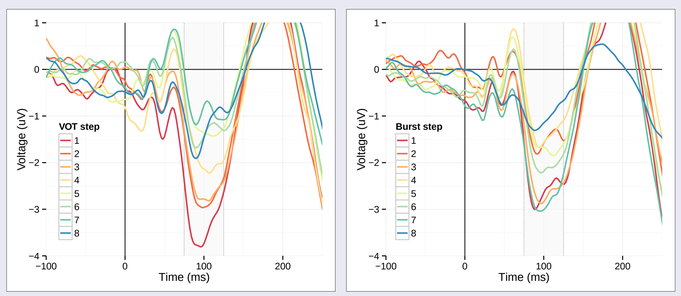
Analysis	Factor	Result
Stops	Place	$\chi^2(2)=3.7$, p=.158
Stops	Voicing	$\chi^2(1)=3.6$, p=.056
Stops	Place × Voicing	$\chi^2(1)=1.4$, p=.506

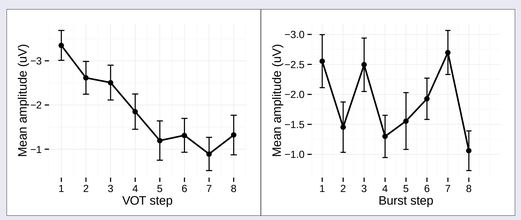
Analysis	Factor	Result
Fricatives	Place	$\chi^2(2)=.18$, p=.915
Fricatives	Voicing	$\chi^2(1)=3.7$, p=.056
Fricatives	Place × Voicing	$\chi^2(1)=5.9$, p=.015

Analysis	Factor	Result
Nasals	Place	$\chi^2(1)=6.0$, p=.015
Affricates	Voicing	$\chi^2(1)=.13$, p=.714
Approximants	Manner	$\chi^2(2)=3.9$, p=.143

RESULTS—EXPERIMENT 2 (VOT and burst continua)

Grand average ERP waveforms (F3,Fz,F4)





Analysis	Factor	Result
Stops	VOT	$\chi^2(7)=36.6$, p<.001
Stops	Burst	$\chi^2(7)=20.6$, p=.004
Stops	VOT × Burst	$\chi^2(49)=49.3$, p=.461

DISCUSSION

Experiment 1

- Differences in N100 amplitude observed for several acoustic and phonological distinctions
- Stops
  - Larger N100 for voiced than for voiceless stops
  - Replicates and extends findings from Toscano et al. (2010)[4] with different manners of articulation
  - Effects of place only observed for voiceless stops
  - Place differences do not reflect encoding based on articulation location; rather, effects appear to map onto spectral shape dimension[5] (/p/→/k/→/t/)
  - Supports N100 encoding based on acoustic cue representations, rather than articulatory-based representations
- Fricatives
  - Effect of voicing similar to stops (larger N100 for voiced consonants); suggests N100 may be sensitive to differences in low-frequency energy common to voicing distinctions across manner of articulation
  - Place × voicing interaction driven by voicing effect for /f,v/
- Nasals
  - Larger N100 for /n/ (alveolar) than for /m/ (bilabial); similar to the place effect for fricatives, and different from direction of place effect for stops

Experiment 2

- Linear effect across VOT continuum, replicating previous work[4]
- Effect of release burst continuum, but non-linear pattern across continuum
- Suggests listeners may not be encoding overall burst spectrum shape, but a more specific cue in the burst (e.g., mean frequency)

Conclusions: The N100 can serve as a useful tool for studying cue encoding at early stages of speech perception across a range of phonological distinctions

ACKNOWLEDGEMENTS/REFERENCES

- We would like to thank Nicole Johnson, Chris Burley, Jackie Coelho, Janice Puder, Nicole Feeley, Caroline Lee, Liz Day, and Gwen Saccocia for assistance with data collection, and Emma Folk for stimulus recording and subject recruitment

- **References:**

1. Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54* 358–368.
2. Pisoni DB, Tash J (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285–290.
3. Miller JL (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271–285.
4. Toscano JC, McMurray B, Dennhardt J, Luck SJ (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21*, 1532–1540.
5. Stevens KN, Blumstein SE (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*, 1358–1368.