3pSC26

The N1 Event-Related Potential Component as an Index of Speech Sound Encoding for Multiple Phonetic Contrasts

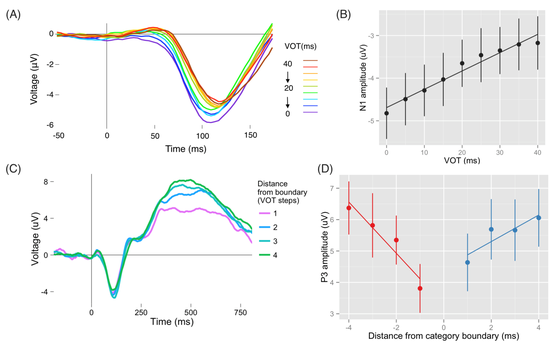
Olivia Pereira (opereira@villanova.edu)
Villanova University

Joe Toscano (joseph.toscano@villanova.edu)
Villanova University

WRAP Lab
http://wraplab.co/

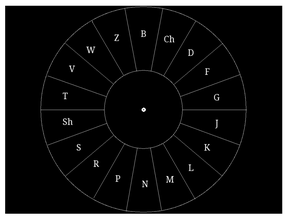
INTRODUCTION

- Early work suggested that fine-grained differences between phonetic cues were discarded early in perception, but recent evidence has refuted this claim [1-3]
- However, it is not clear how the perceptual system encodes specific phonetic cues at early stages
- Previous work has used the ERP technique to address this: [4]
  - Auditory N100 component varies linearly with differences along VOT continua
  - P300 varies as a function of phonological category, while retaining sensitivity to within-category differences
- We aimed determine how N1 varies for other cues across a wide range of consonant distinctions



METHOD

- Subjects heard word-initial minimal pairs spanning 18 consonants (/b,ʧ,d,f,g,ʤ,k,l,m,n,p,r,s,ʃ,t,v,w,z/) while brain activity was recorded
- Listeners indicated the sound that each word began with by clicking corresponding letter(s) on a display

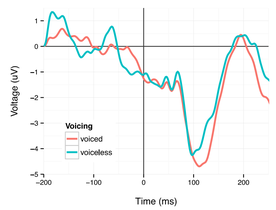


- EEG recorded using a 64-ch Brain Vision actiCHamp system
- Electrodes placed at 10-20 sites (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, T7, T8, O1, Oz, O2, CP1, CP2, FC1, FC2, A1, A2, FP1, FP2)
- EOG recorded via electrodes at FP1 and FP2 (vertical) and electrodes lateral to the external canthi of each eye (horizontal)
- Electrode impedences <5 kΩ
- Data referenced online to the left mastoid (A1), and referenced offline to average mastoid
- EEG digitized at a sampling rate of 500 Hz and recorded in DC mode; offline bandpass filter from 0.1 to 125 Hz was applied in ERPLAB (Butterworth filter w/ 36 dB/octave roll-off)
- Mean ambient room temperature in chamber: 19.7°C
  Mean relative humidity: 29.6%.

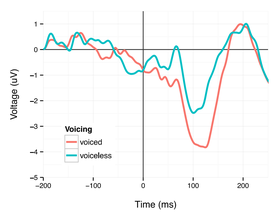
RESULTS

Stop Consonants

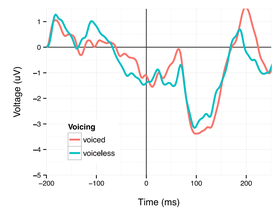
Bilabial (/b/ vs. /p/)



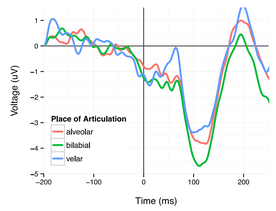
Alveolar (/d/ vs. /t/)



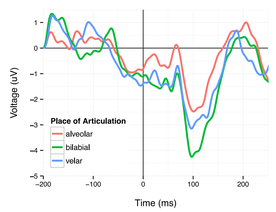
Velar (/g/ vs. /k/)



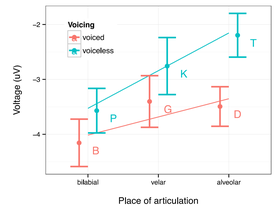
Voiced (/b/ vs. /g/ vs. /d/)



Voiceless (/p/ vs. /k/ vs. /t/)

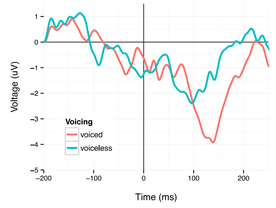


Mean N1 Amplitude

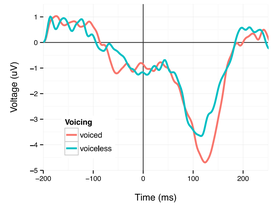


Fricatives

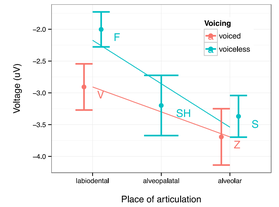
Labiodental (/v/ vs. /f/)



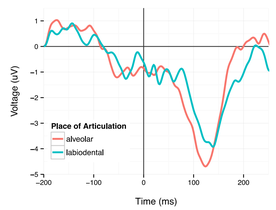
Alveolar (/z/ vs. /s/)



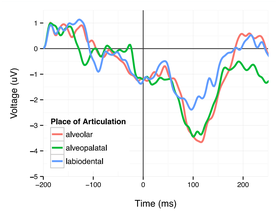
Mean N1 Amplitude



Voiced (/z/ vs. /v/)

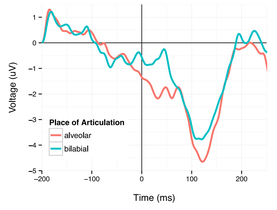


Voiceless (/f/ vs. /ʃ/ vs. /s/)

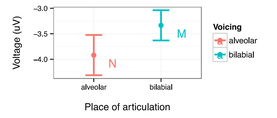


Nasals

Voiced (/m/ vs. /n/)

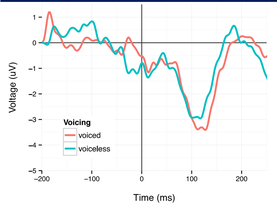


Mean N1 Amplitude

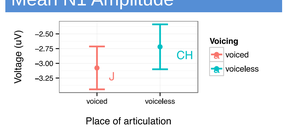


Affricates

Alveopalatal (/ʤ/ vs. /ʧ/)

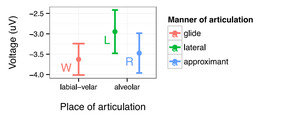


Mean N1 Amplitude

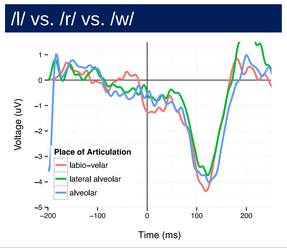


Approximants

Mean N1 Amplitude



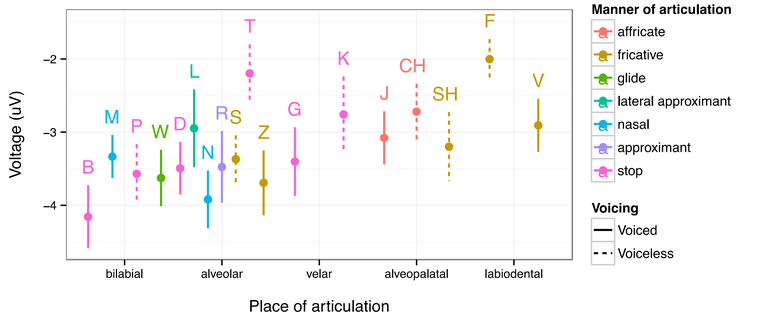
/l/ vs. /r/ vs. /w/



N1 Amplitude—All Consonants

Analysis	Factor	Result
Stops	**Place of articulation**	**$\chi^2(2)=6.56$, p=0.038 ***
Stops	*Voicing*	***$\chi^2(1)=3.10$, p=0.078*** .
Stops	Place × Voicing	$\chi^2(2)=0.85$, p=0.654
Fricatives	**Place of articulation**	**$\chi^2(2)=7.82$, p=0.020 ***
Fricatives	Voicing	$\chi^2(1)=1.12$, p=0.291
Fricatives	Place × Voicing	$\chi^2(2)=0.62$, p=0.432
Nasals	Place of articulation	$\chi^2(1)=1.95$, p=0.16
Affricates	Voicing	$\chi^2(1)=0.41$, p=0.52
Approximants	Manner of articulation	$\chi^2(2)=1.18$, p=0.55

Mean N1 Amplitude



DISCUSSION

- Differences in N1 amplitude observed as a function of voicing (stop consonants) and place of articulation (stops and fricatives)
- **Voicing**
  - Larger N1 for voiced than voiceless consonants
  - Replicates and extends findings from Toscano et al. (2010) [4]
  - N1 may index degree of low-frequency voicing energy that occurs consistently across consonants with different manners of articulation
- **Place of articulation**
  - Pattern of responses varies depending on manner of articulation
  - For fricatives:
    - N1 amplitude larger for alveolar than labiodental sounds
    - No clear distinction between /s/ and /ʃ/ implies that N1 may not be sensitive to cues for this phonetic contrast
  - For stops:
    - Differences do not reflect encoding based on location of articulatory constriction
    - Instead, effects map onto acoustic differences in burst frequency (/p/ → /k/ → /t/)
    - Supports the N1 as a measure of acoustic cue encoding, rather than an articulatory-based representations
    - Follow-up work will explore these results in detail by varying sounds along VOT (voicing) and burst frequency (place) continua
- **Conclusion:** Overall, results suggest that the N1 can serve as a useful tool for studying cue encoding at early stage of perception across a range of speech distinctions

ACKNOWLEDGEMENTS & REFERENCES

- We would like to thank Nicole Johnson, Chris Burley, and Jackie Coelho for assistance with data collection, and Emma Folk for stimulus recording and subject recruitment.

- **References**

- Liberman AM, Harris KS, Hoffman HS, Griffith (1957). *Journal of Experimental Psychology, 54*, 358-368.
- Pisoni DB, Tash J (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285-290.
- Miller JL (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271-285.
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (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.