Human speech cortex long-term recordings [5]: Formant frequency analyses

J. S. BRUMBERG[1], D. S. ANDREASEN[2], J. L. BARTELS[2], F. H. GUENTHER[1], P. R. KENNEDY[2], S. A. SIEBERT[2], A. B. SCHWARTZ[3], M. VELLISTE[3], E. J. WRIGHT[2]

[1]Cognitive and Neural Systems, Boston University, Boston, MA. [2]Neural Signals Inc, Atlanta, GA. [3]Motorlab, University of Pittsburgh, Pittsburgh, PA

517.17/VV3

Introduction

The neural correlates of speech production have been studied in humans using fMRI and other brain imaging techniques. However, extracellular electrical recordings are not often used and offer improved characterizations of single neurons in regions involved with the production and perception of speech. In addition, investigations of single neurons in the motor and premotor cortices provide a means by which effective movement restoration neural prostheses can be designed and implemented.

A patient with locked-in syndrome, resulting from a brain stem stroke, was implanted with the neurotrophic electrode [1] in the speech-motor related left premotor cortex. Single units were isolated from the raw spike waveform according to standard manual cluster cutting techniques. Details concerning the electrode, implantation procedure, and cluster cutting are described elsewhere [see poster 728.14/OO13].

The patient participated in a vowel repetition task in which he listened to acoustic vowel stimuli and attempted to reproduce the sounds. Action potential occurrence times of 41 premotor cortex neurons were analyzed with respect to salient acoustic characteristics of the vowel stimuli, specifically formant frequencies, during both passive listening and active production.

Methods

Stimuli

Three steady state vowels were used: /A/, /IY/ and /OO/ (as in hot, heat, hoot) of 500ms duration

Task

The task was divided into two conditions: speech perception and production

Each stimulus was presented acoustically to the subject (PERCEPTION) followed by a silence period. The patient was asked to repeat the stimulus as he heard it after the presentation of a GO signal (PRODUCTION). On some trials no acoustic stimulus was presented. In this case the patient was asked to make no productions until the next trial began.

Single Unit Analysis

Raw DC waveforms were bandpass filtered with passband 300-6000 Hz. Threshold crossings were used to identify putative action potentials. This analysis uses the spike clusters discussed in poster 728.14/OO13.

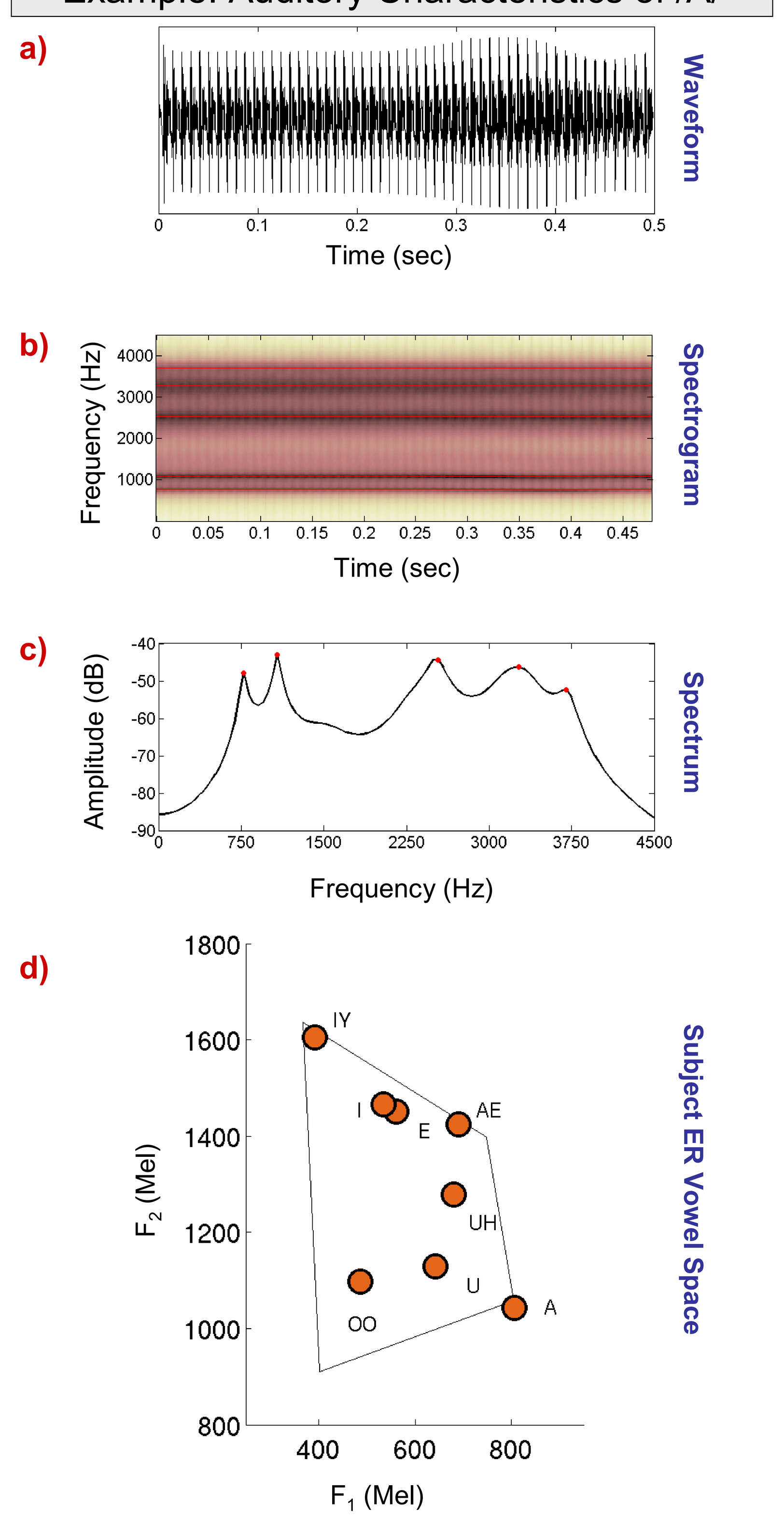
Spike Train Analysis

1. Single unit spike trains were analyzed in two time regions: 0-750 ms from the stimulus onset (PERCEPTION) and 0-2 sec from the GO signal (PRODUCTION). Vowel responses were predicted using optimal linear decoding [3] of the ensemble firing rate.
   - Firing rates: peri-event time histograms computed with 25 ms bins rates log-transformed and normalized by overall activity.

2. Vowel formant frequencies are used as optimal linear decoding response variables. Formants are peaks in the acoustic spectrum, representing the most important acoustic information for vowel perception (see panel on right). Vowels can be characterized by their first two formant frequencies, F_1 and F_2, which are related to vocal tract articulator positions.
   - Formant frequencies for each vowel were chosen to match the average formant frequencies utilized by the patient's father (referred to as subject ER herein).
   - Formant frequencies are analyzed in Mel units rather than Hz; they provide a better description of human acoustic perceptual space.

3. Evaluation of predicted formant frequencies was accomplished using linear discriminant analysis (LDA):
   a. Predicted formant frequencies were classified as specific vowels
   b. Single time bin predicted vowel classification confidence (based on formant frequency estimates) was evaluated using the ratio of the maximum posterior probability Pr(group j| obs i) to the next highest probability [4].
   c. Bin-wise strength was computed for trials of a particular vowel as follows. For each time bin, the percentage of trials that were classified as a particular vowel were calculated and plotted as a function of trial bin for each vowel [4].

Example: Auditory Characteristics of /A/



Formant frequencies plotted in red in spectrogram (b) and spectrum (c) plots. Subject ER's vowel space plotted over the standard vowel quadrilateral (d).

References

1. Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K. and Goldwaithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8(2):198-202.
2. Siebert, S. A., Andreasen, D. S., Bartels, J. L., Brumberg, J. S., Guenther, F. H. Kennedy, P. R. and Wright, E. J. (2007). Human speech cortex long-term recordings [1]: Spike sorting and noise reduction. Abstract No. 728.14/OO13, 2007 Society for Neuroscience Meeting.
3. Warland, D. K., Reinagel, P. and Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78:2336-2350.
4. Averbeck, B. E., Chafee, M. V., Crowe, D. A. and Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99(20):13172-13177.
5. Breiman, L., Friedman, J. Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*, Belmont, CA: Wadsworth International Group.
6. Georgopoulos, A. P., Schwartz, A. B. and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771):1416-1419.

Contact: Jon Brumberg brumberg@cns.bu.edu

Results

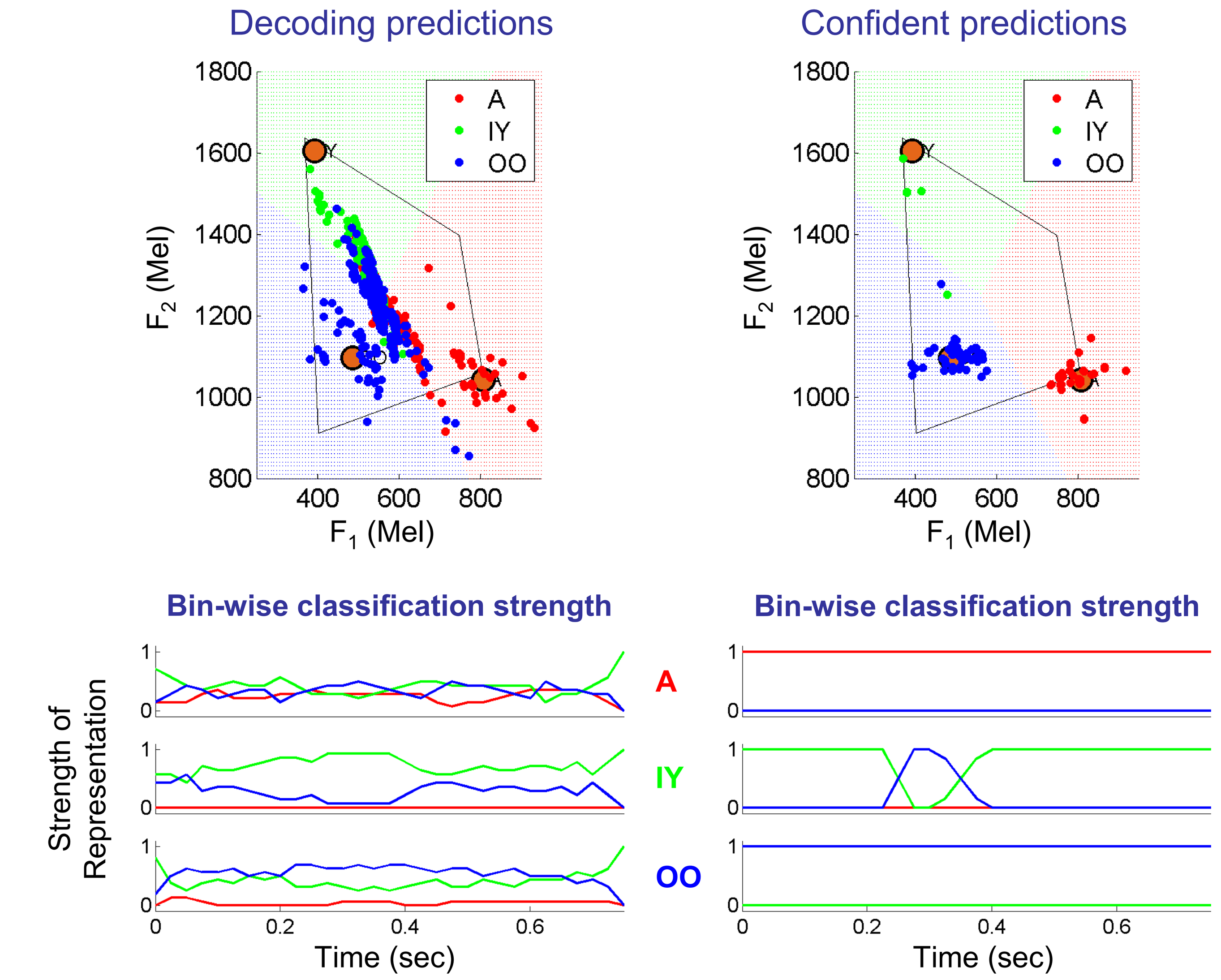
Spike Train Analysis

1. Optimal linear decoding was applied to the spike train data using two regression techniques: least squares and regression trees [5]
   - Significantly active neurons determined via stepwise least squares regression
     - Regression coefficients represent neurons; coefficients were iteratively added and removed based on a statistical threshold test (P<0.05).
   - N=(13,14) F1, F2 neurons during PERCEPTION; N=(16,12) F1, F2 neurons during PRODUCTION
   - N=(5, 5) F1, F2 neurons active during both PERCEPTION and PRODUCTION – **MIRROR NEURONS?**
     - 3 neurons active for similar directions of F1 in the two conditions, 3 neurons for similar directions of F2
2. Vowel classification decision regions were determined via LDA classification of vowels from predicted formant frequencies.

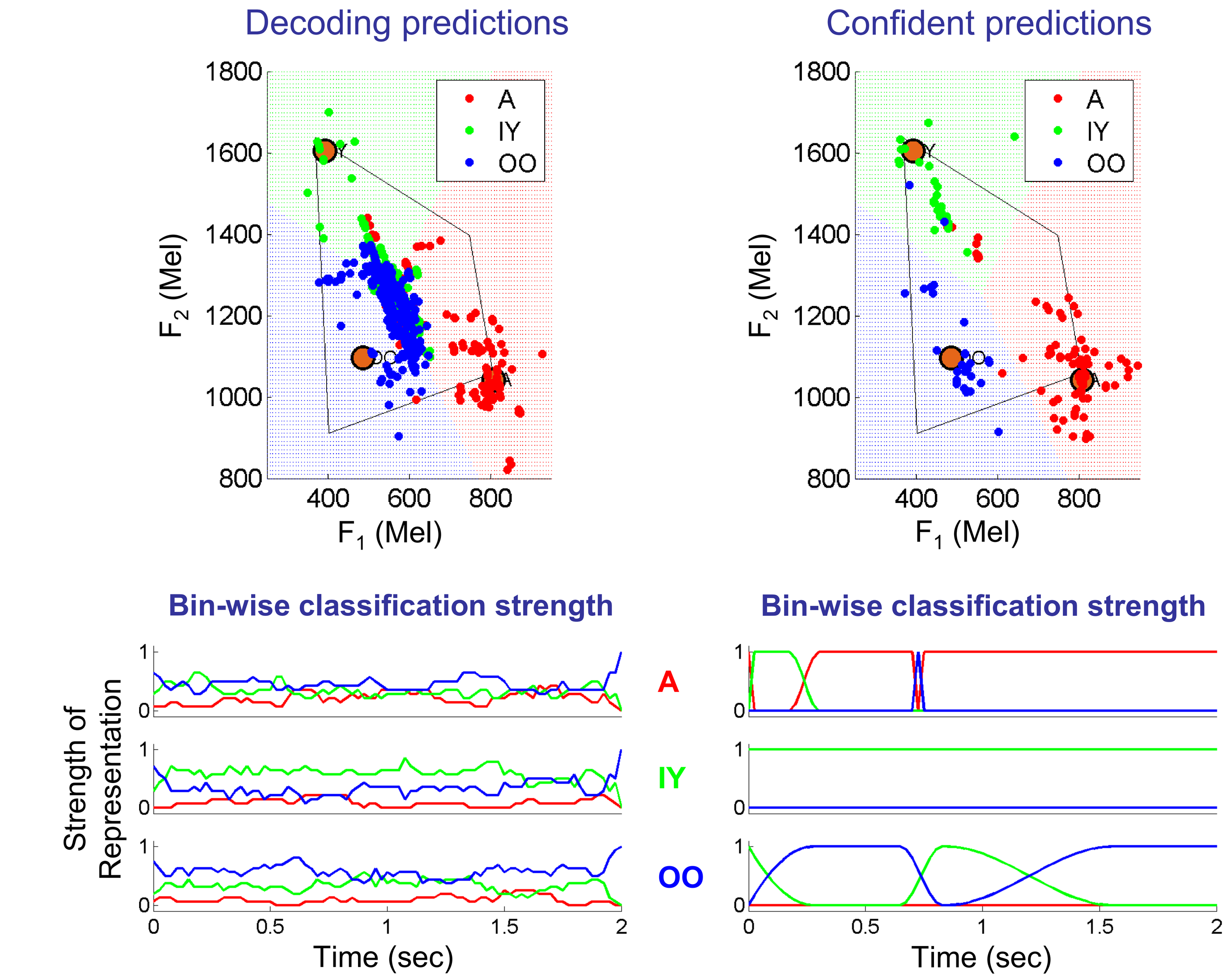
Panels A and B

- Each data point represents the predicted formant frequencies from the firing rate in a single time bin
- Decoding prediction colors indicate ground truth vowel membership
- Background grid colors illustrate LDA decision boundaries for vowel classification
- Confident predictions are predictions from the optimal linear decoder retrained using data points with above threshold classification confidence
  - Confidence threshold chosen based on sorted confidences for a given time bin ("knee" of exponential fit to data)
  - These confident predictions identify times when subject made a response and/or times when neuron was active
- Bin-wise classification strength indicates average probability of choosing a particular vowel at every trial time point
  - Computed for both the full ensemble firing rate data and for the subset of confident predictions

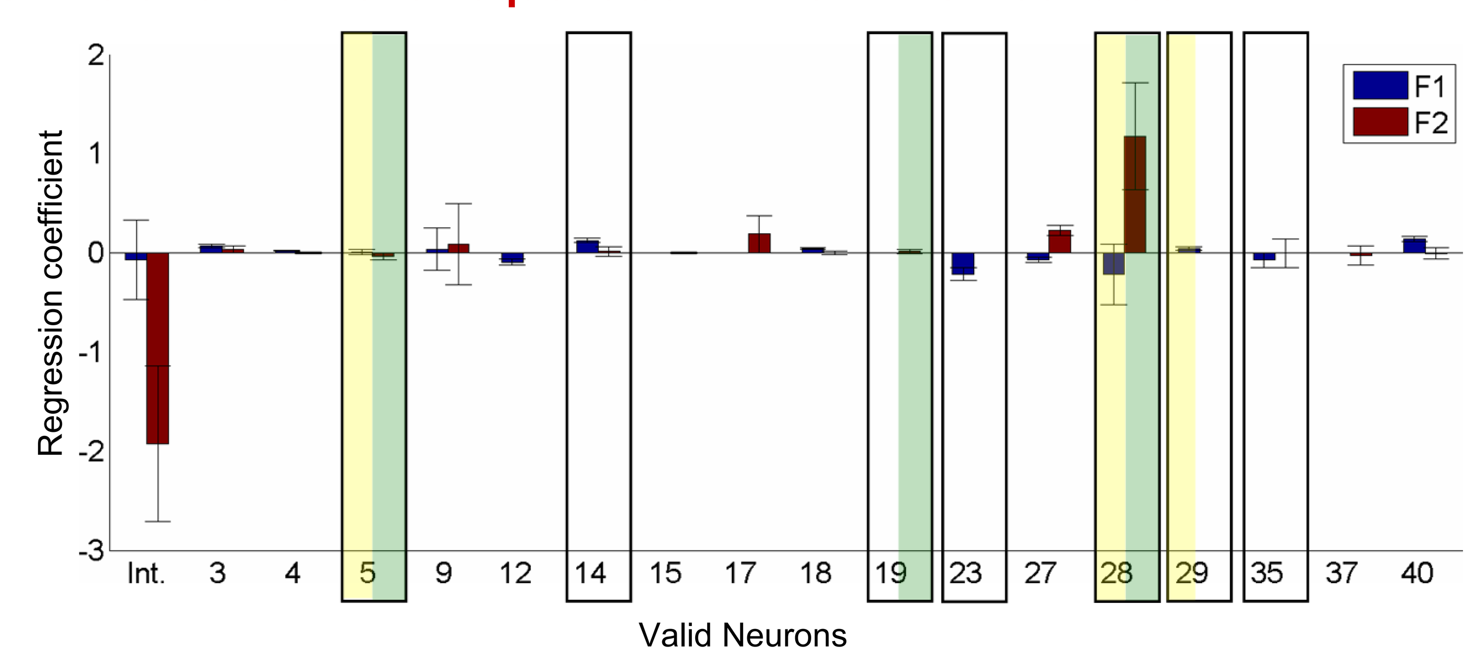
A. PERCEPTION condition



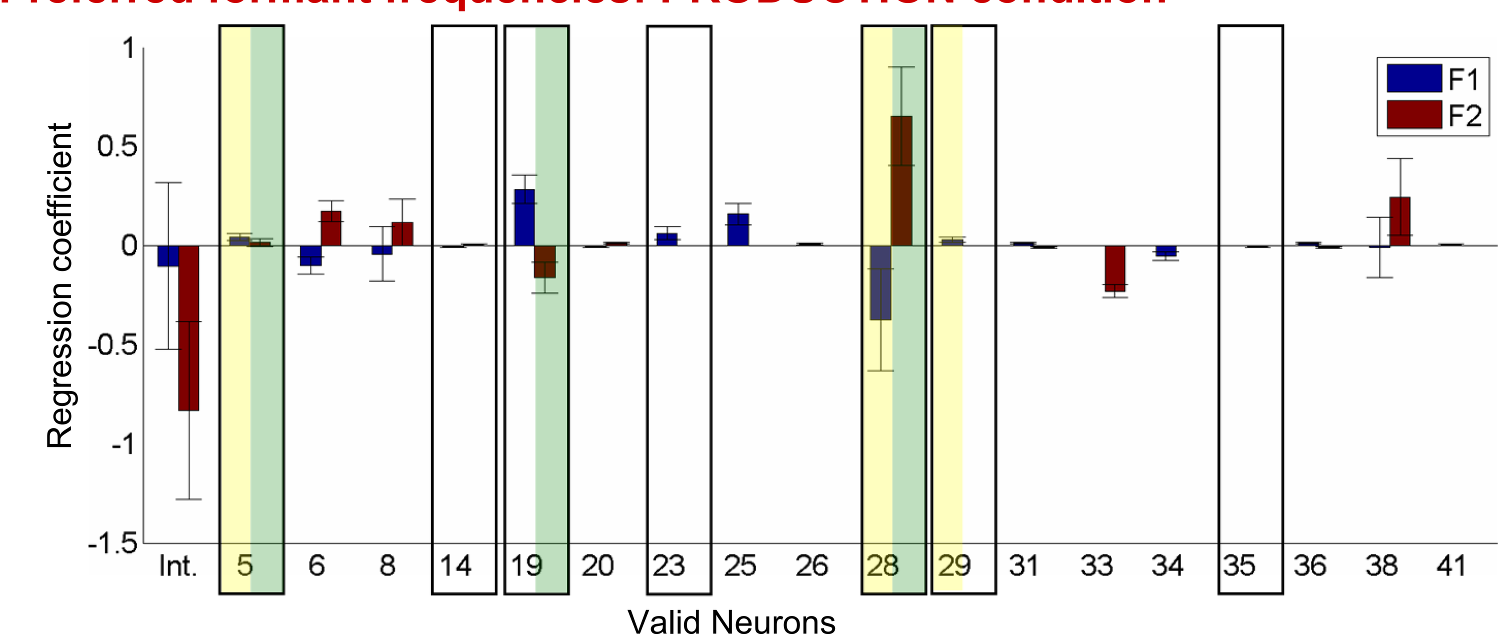
B. PRODUCTION condition



C. Preferred formant frequencies: PERCEPTION condition



D. Preferred formant frequencies: PRODUCTION condition



Above: Regression coefficients for decoding during PERCEPTION (C) and PRODUCTION (D) conditions (including intercept, for interpretation of coefficient magnitudes)
- Direction of bars indicates preference for *high* (positive values) or *low* (negative values) formant frequencies with F1 in blue and F2 in red
- Boxes indicate neurons active in both production and perception conditions, possible mirror neurons
  - Yellow shaded regions are neurons with similar F1 preference
  - Green shaded regions are neurons with similar F2 preference
- Error bars are 95% confidence intervals of coefficient estimates

PERCEPTION:
- F1 intercept: 620 Mel
- F2 intercept: 1115 Mel

PRODUCTION:
- F1 intercept: 640 Mel
- F2 intercept: 1150 Mel

PERCEPTION decoding coefficients have higher variance than PRODUCTION coefficients

⟶ **Single units are more finely tuned to production than perception**

Conclusions

1. In both production and perception conditions, populations of cells show significant main effects of first and second formant frequency
2. An overlapping population of cells show modulation of activity during both production and perception
   1. Some of these cells appear to code for the same regions (high vs. low) of formant frequencies across conditions
   2. Suggests the possibility of mirror neuron activity
3. Future analysis will include stimuli designed to examine neural activity with respect to directions, rather than positions, in the formant plane as motor cortex neurons have been found to be sensitive to movement direction [6]

This study is currently active; new encoding and decoding methods are being tested for:
1. Their capability to characterize the firing preferences of premotor cortex neurons during speech production and perception
2. Performance in a neural prosthesis for speech

This research is supported in part by the National Institutes of Health (NIH R01 DC-007683), the National Science Foundation (NSF SBE-0354378), the National Institute of Deafness and Other Communication Disorders (NIDCD 2 R44 DC007050-02). Work conforms to IACUC and FDA guidelines. Conflict of interest for PRK and DA.