Estimating speech sound categorization from electrophysiological responses


Daniel McCloy


Adrian KC Lee

Institute for Learning and Brain Sciences
University of Washington

Motivation

- **LANGUAGE SCIENCE:** new view of mental models of speech
- **HEALTH:** possibly useful in assessments of CIs, dyslexia, etc.

Challenges

- Only very low-frequency stimulus modulations are preserved in the EEG signal
- Tonotopy is hard to resolve at scalp
- Natural stimuli (speech) are not always easy to manipulate
- Dimensions of stimulus variation maybe not orthogonal

Our (novel?) approach

- Lots of stimulus types varying in many ways
- Use supervised machine learning to find regularities in EEG responses

Example: Consonant Classification

- **STIMULI**: English & foreign consonant-vowel (CV) syllables; variable consonant, vowel always [ɑ]
- **TRAINING SET (ENGLISH)**: 2 talkers (♂/♀) × 3 recordings × 23 consonants × 20 presentations = 2760 trials
- **TEST SET (ENGLISH)**: 2 new talkers (♂/♀) × 1 recording × 23 consonants x 20 presentations = 920 trials
- **TEST SETS (DUTCH/HUNGARIAN/HINDI/SWAHILI)**: 1 talker {♀/♀/♂/♂} × {18/25/30/30} consonants × 20 presentations = {360/500/600/600} trials
- **RECORDING**: 32-channel BrainVision, left earlobe reference, 1000 Hz sampling rate
- **PREPROCESSING**: bandpass 1-40 Hz, downsample to 100 Hz, align epochs on boundary between C and V, apply denoising source separation[1,2] (DSS), remove time domain autocorrelation with PCA (retains 20 "samples"), use first 4 DSS components
- **SUPERVISED LEARNING**: label all syllables with phonological distinctive features from PHOIBLE[3] database, train binary classifier (support vector machine with radial basis function) for each distinctive feature (5-fold cross-validation + grid search), handle class imbalance by setting threshold to equalize error rate (false positive rate = false negative rate)
- **EVALUATION**: apply classifiers to test data, estimate "what listener heard" as maximum joint probability of classifiers: $P(\text{“d”}) = P(+\text{VOICED}) \times P(+\text{CORONAL}) \times \ldots \times P(-\text{SONORANT})$

Results

- Fairly consistent results across-subj. for English; need more data for other langs. (more tokens or more talkers?)

Future directions

- Vary both consonants and vowels
- More languages / speech sound types (airstream and phonation contrasts, tone)
- Increase SNR: more data, different classifier strategies, simultaneous MEG + EEG experiments
- Use unsupervised learning to derive optimal, perceptually-based phonological distinctive features

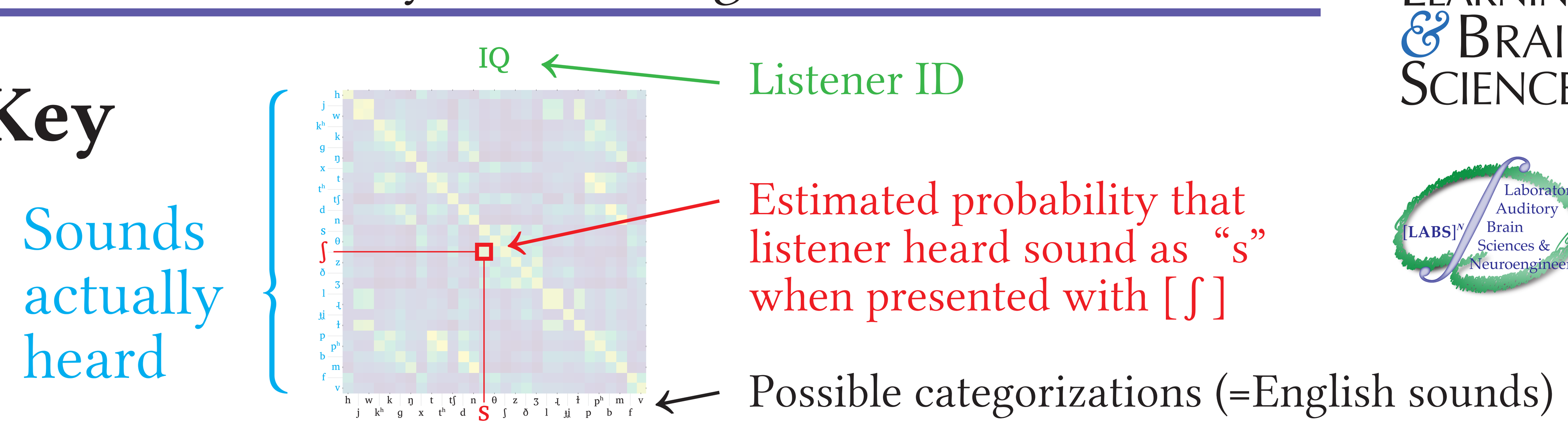
Other applications of this method

- Classify cells based on spiking pattern?
- Diagnostic use for cochlear implants, language impairments?

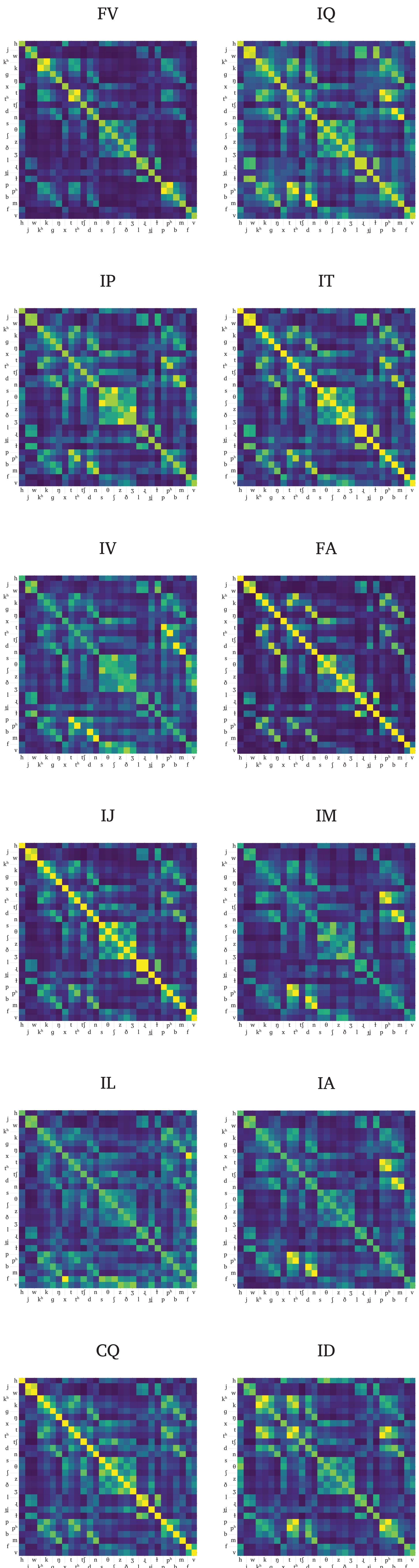
Acknowledgments

NSF Jelenik Speech and Language Technology Workshop 2015, Probabilistic Transcription Team; Majid Mirbagheri; Nick Foti; Eric Larson; Mark Hasegawa-Johnson; Preethi Jyothi.

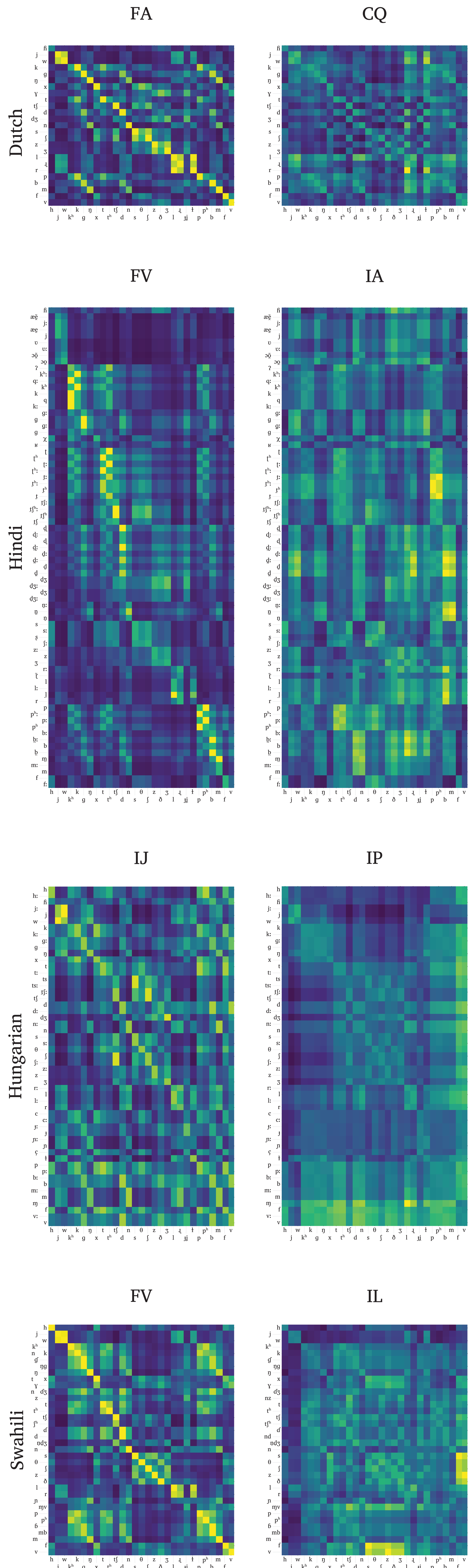
Key



All listeners' computed confusion matrices (English test stimuli)



Best/worst computed confusion matrices (Foreign test stimuli)



References

[1] J. Särelä and H. Valpola, "Denoising source separation," *J. Mach. Learn. Res.*, vol. 6, pp. 233–272, 2005.

[2] A. de Cheveigné and J. Z. Simon, "Denoising based on spatial filtering," *J. Neurosci. Methods*, vol. 171, no. 2, pp. 331–339, 2008.

[3] S. Moran, D. McCloy, and R. Wright (eds). *PHOIBLE: Phonetics Information Base and Lexicon Online.* Munich: Max Planck Digital Library, 2013.