

Towards a biological theory of phonetic perception

Q

Can we build a theory of phonetic perception from the ground up using neurological primitives?

A

so far, so good

Q

Why would this be desirable?

A

- Many theories at the ‘computational level’ (exemplar theory, Bayesian models, etc...) have no obvious way of linking up with cell biology
- Maybe if we start from the bottom, a good formal characterization of the basic biology will suggest (or force upon us) a theory at the ‘computational level’

How this will work

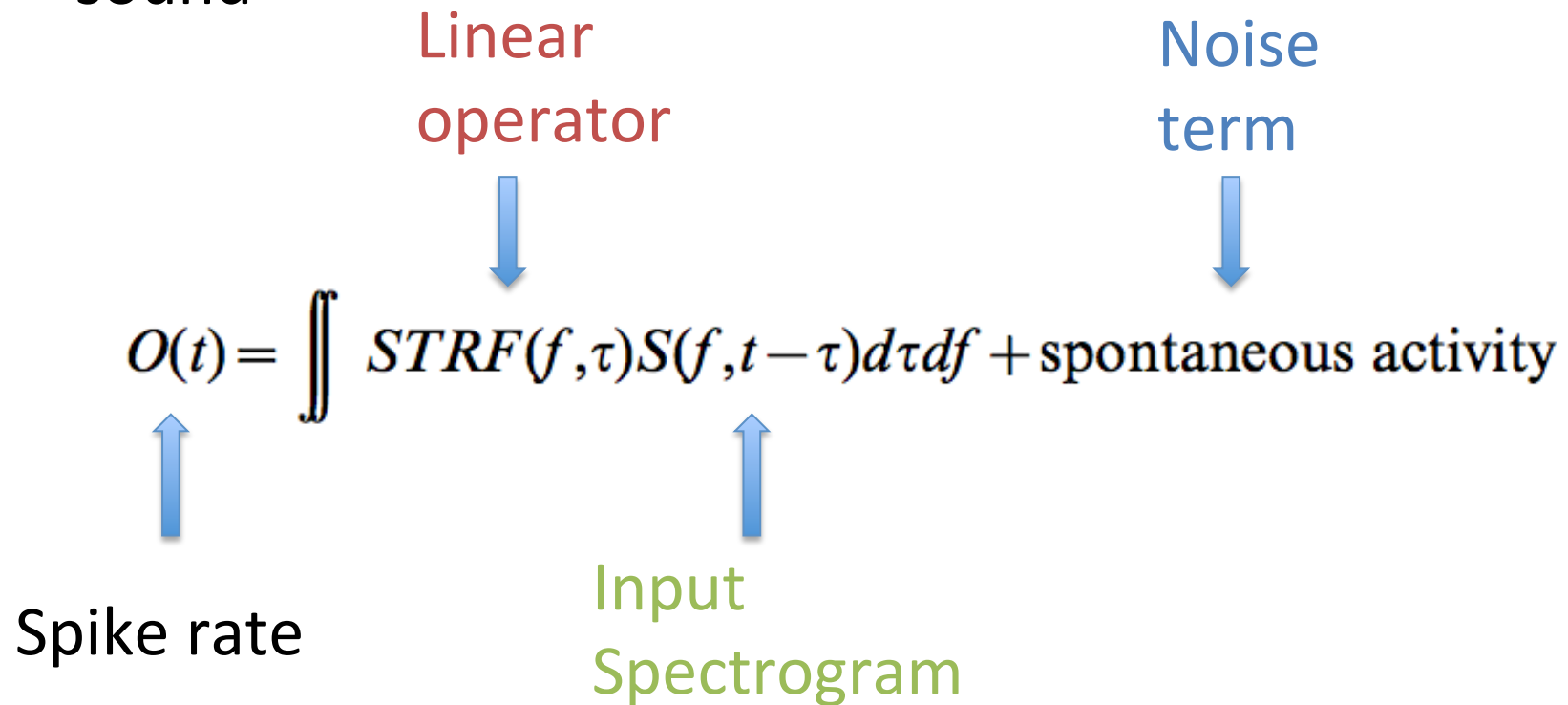
- **Fact:** The preferred stimuli of many neurons in higher auditory areas are complex sounds
- **Fact:** The behavior of these neurons can be modeled as linear operators
- **Hypothesis:** Phonetic categories are an ensemble of specific neurons with specific spectrotemporal preferences which span the space of the sounds of that category
- **Hypothesis:** The characteristics of phonetic perception are a reflex of the structure of these operators

outline

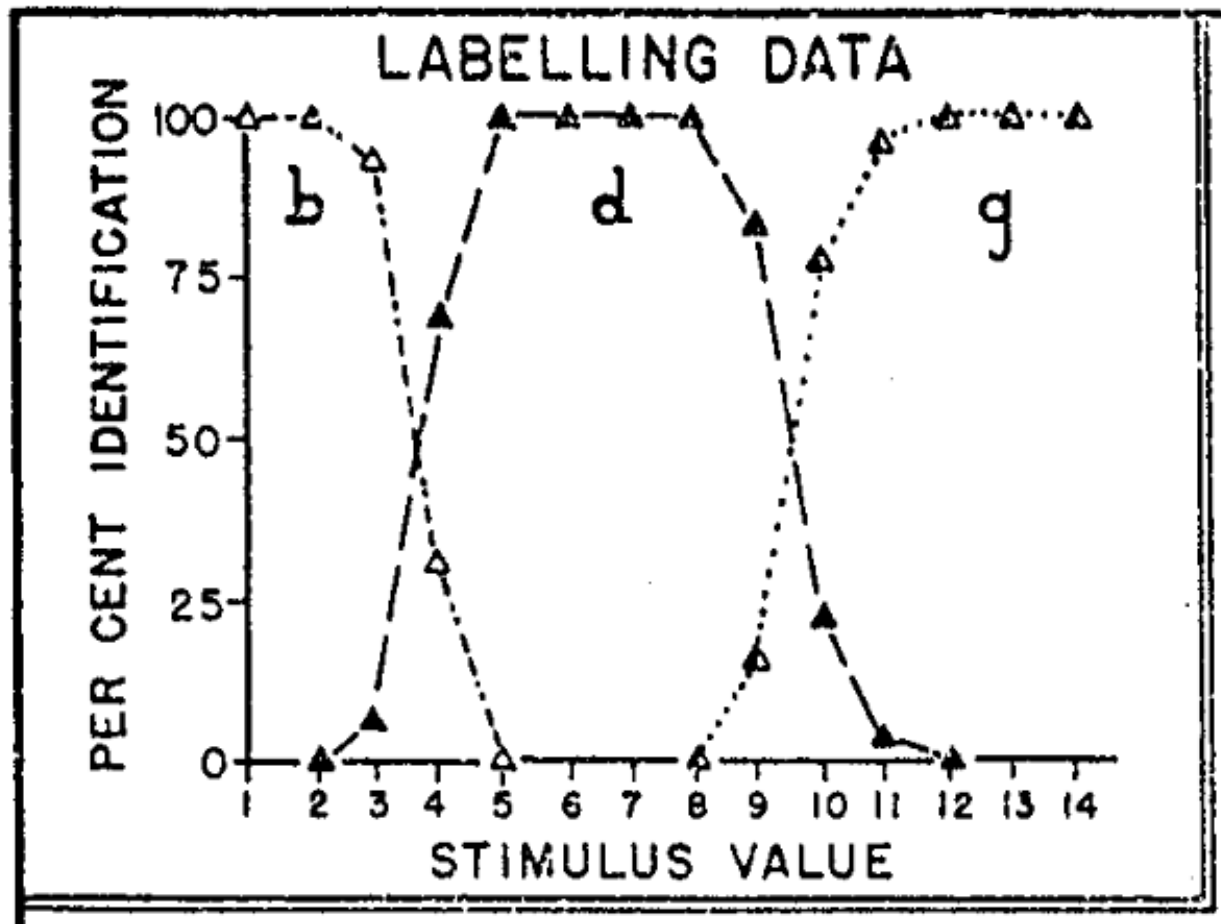
- desiderata
- neurons
- linear operators
- simulation
- discussion

desiderata: building blocks

A (well-behaved) mathematical characterization of the behavior of auditory neurons in response to sound



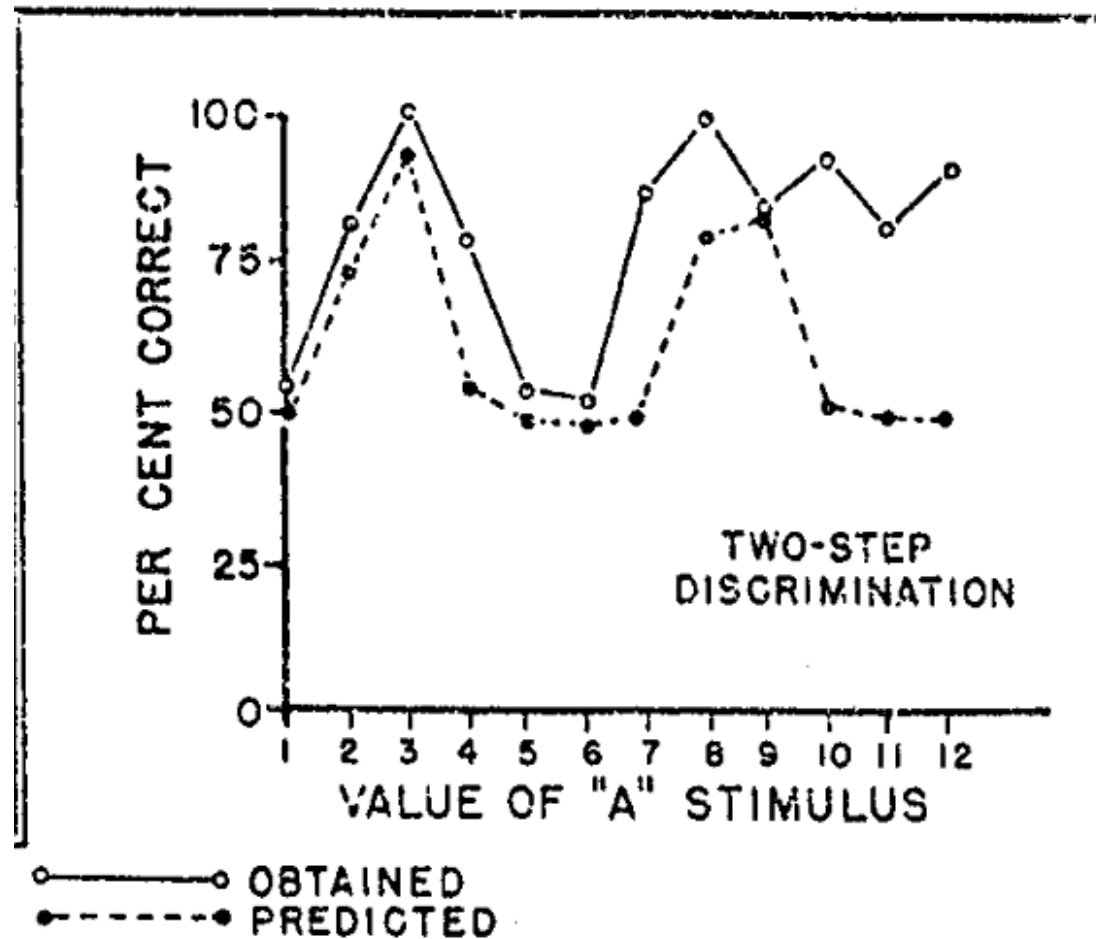
desiderata nonlinear identification functions



Liberman et al, 1957

desiderata

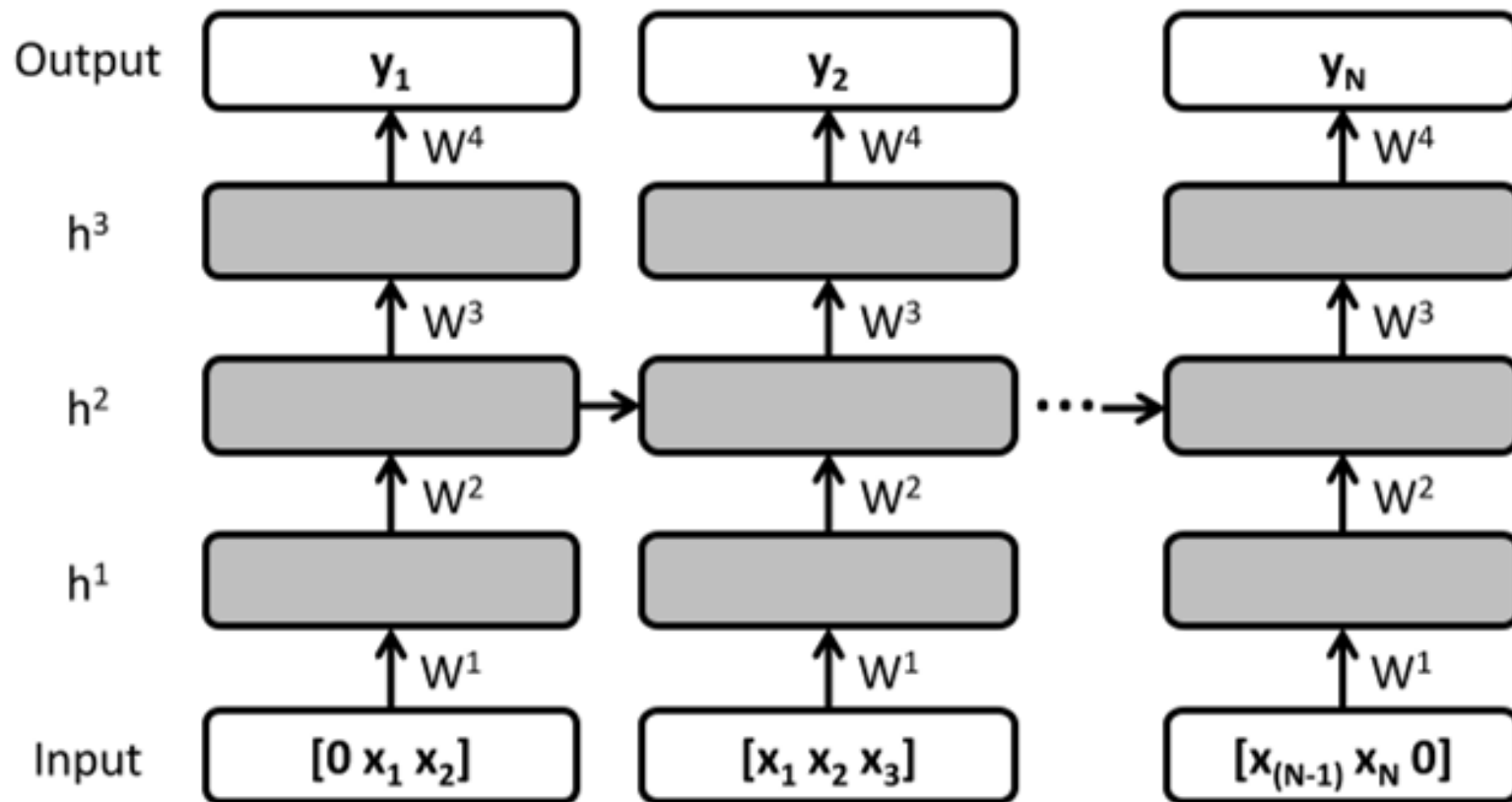
nonlinear discrimination functions



Liberman et al, 1957

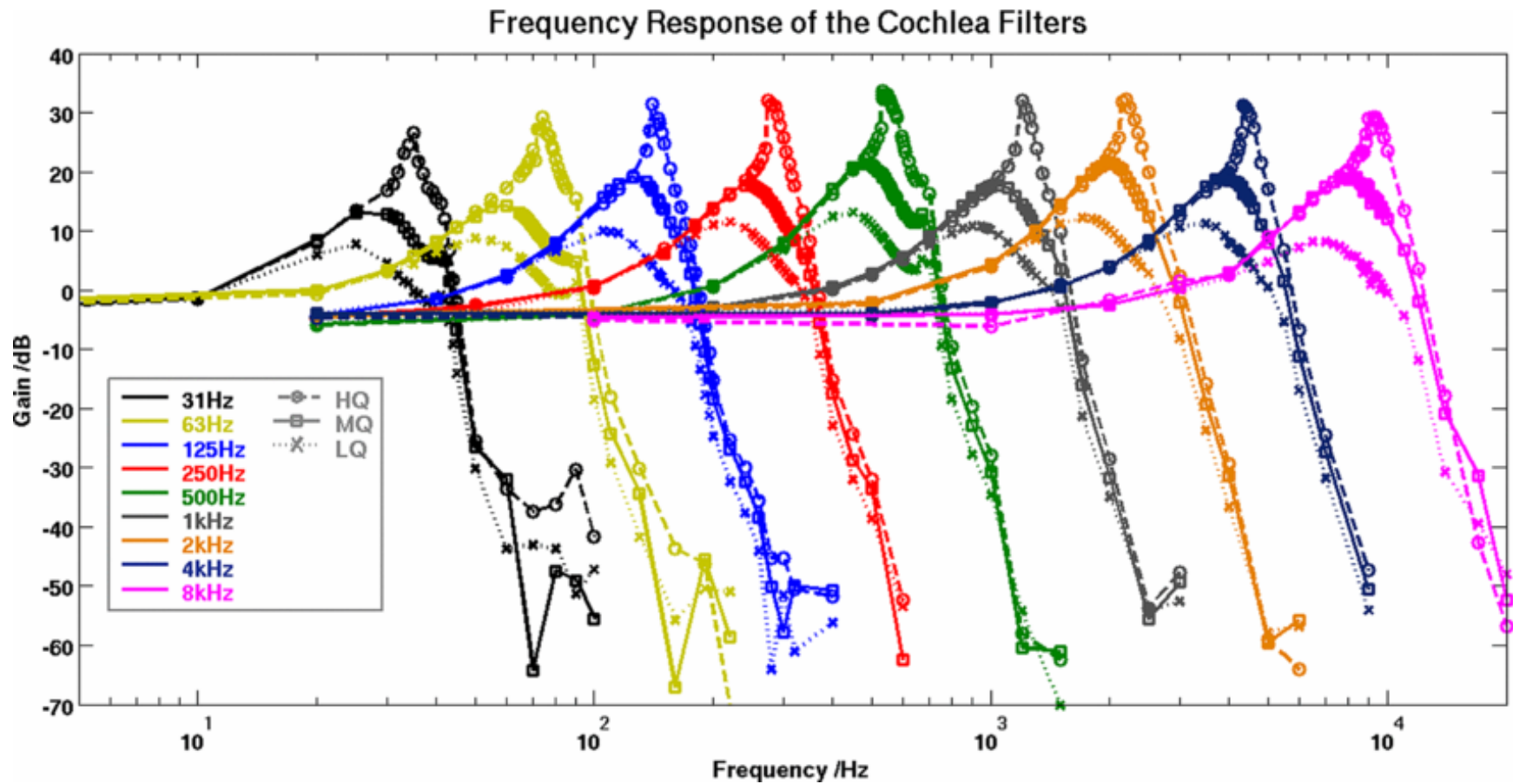
anti-desiderata

acoustic 'features' fed into black boxes



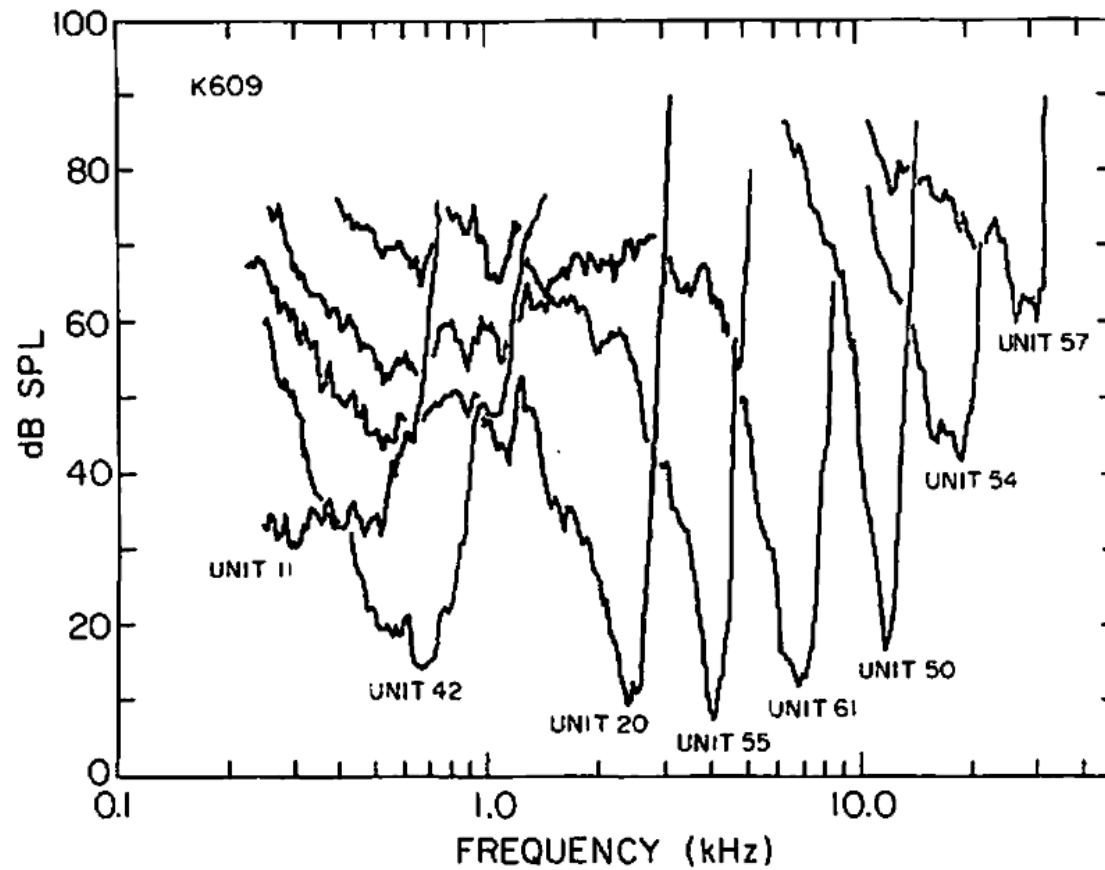
neurons

Cochlear filters



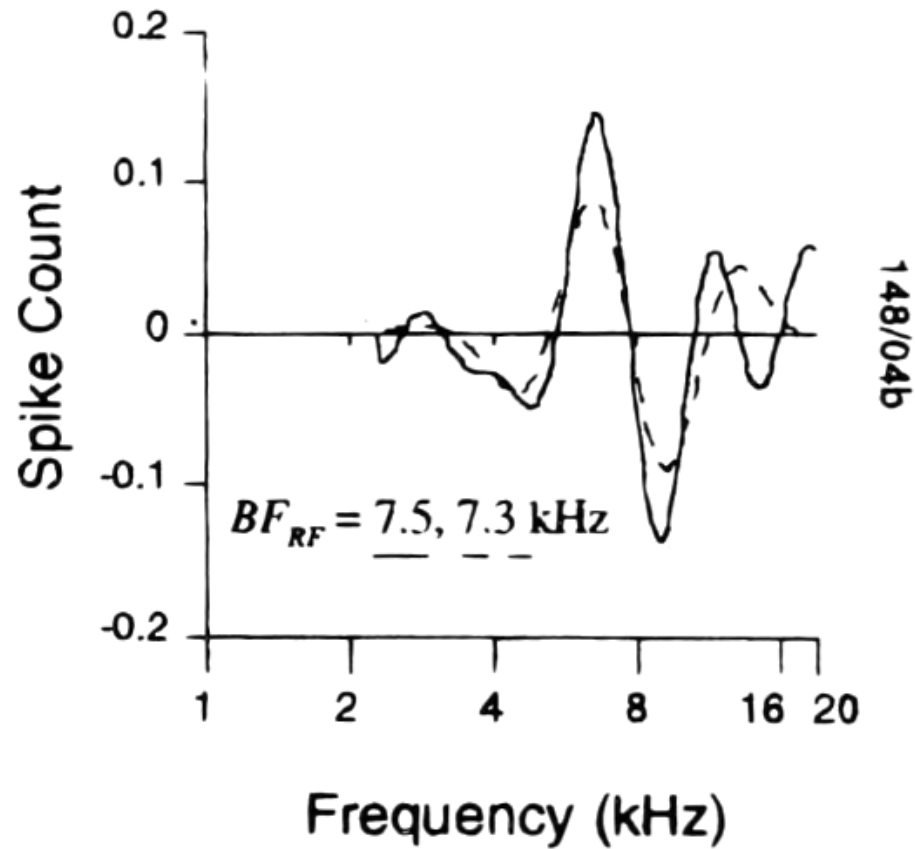
Wang et al, 2014

Auditory nerve (cat)



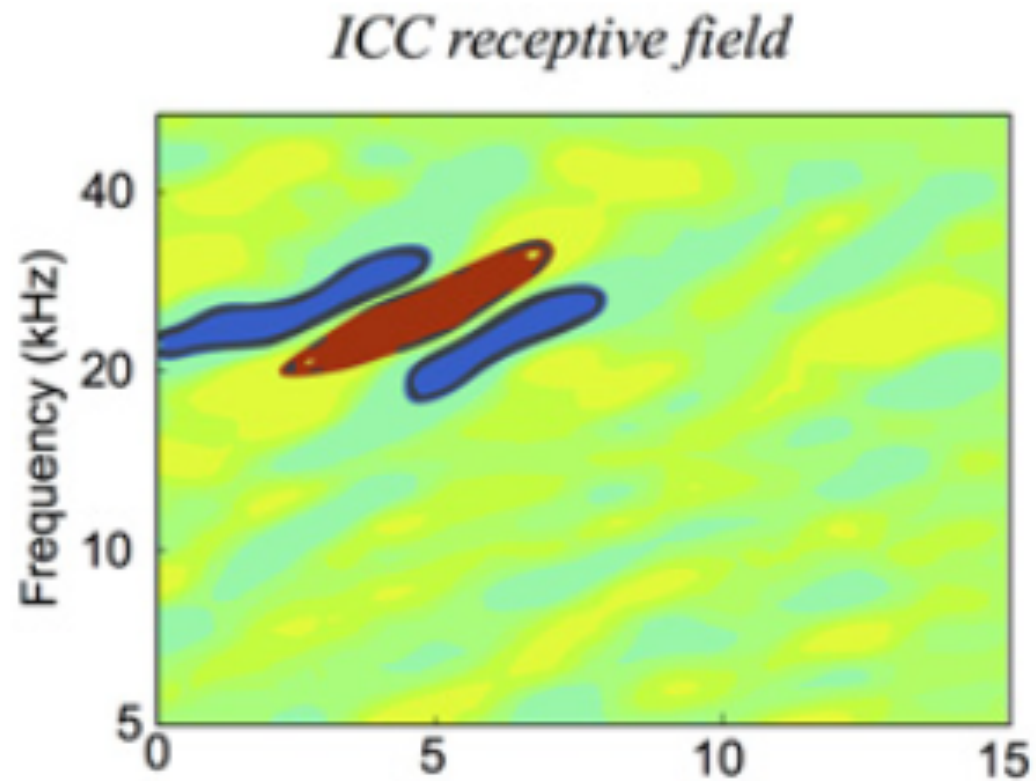
Kiang, 1975

A1 (ferret)



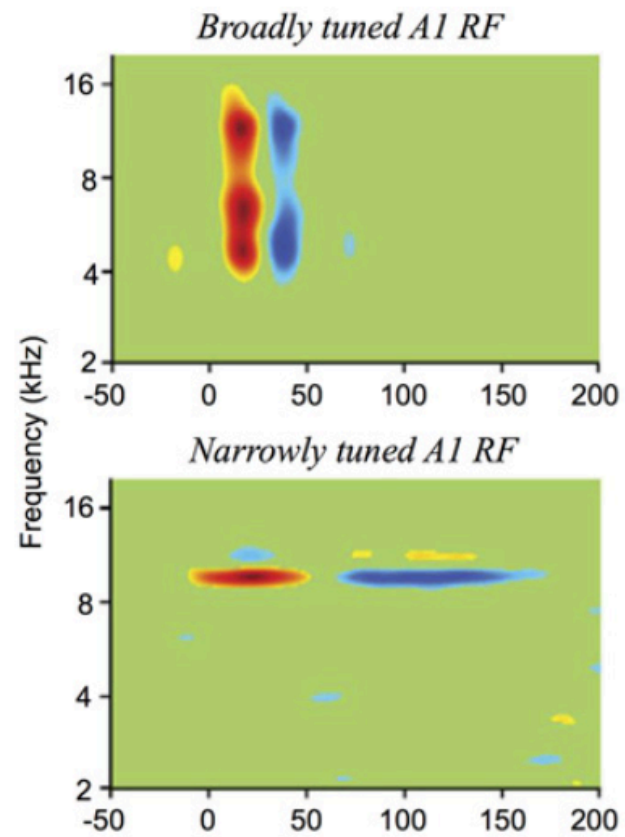
Shamma et al, 1995

ICC (bat)

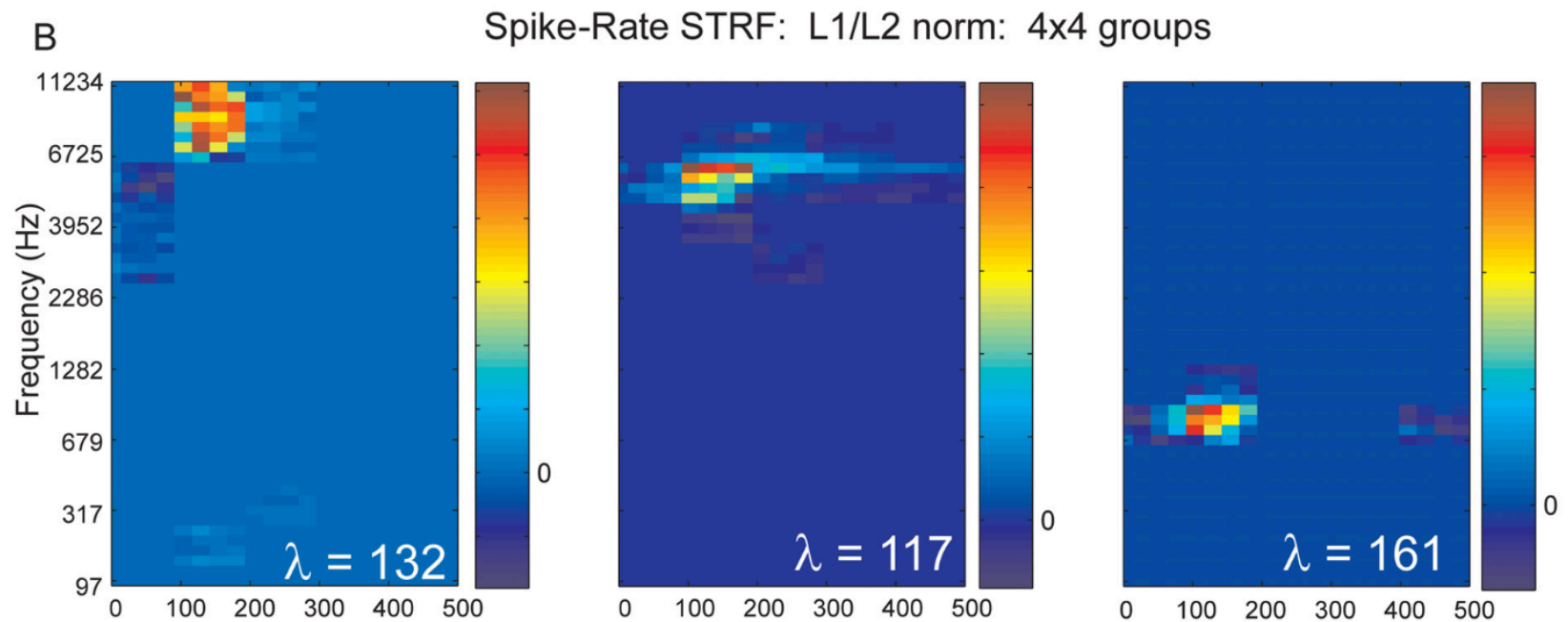


Andoni et al, 2007

A1 (cat)



HG (human, multiunit)



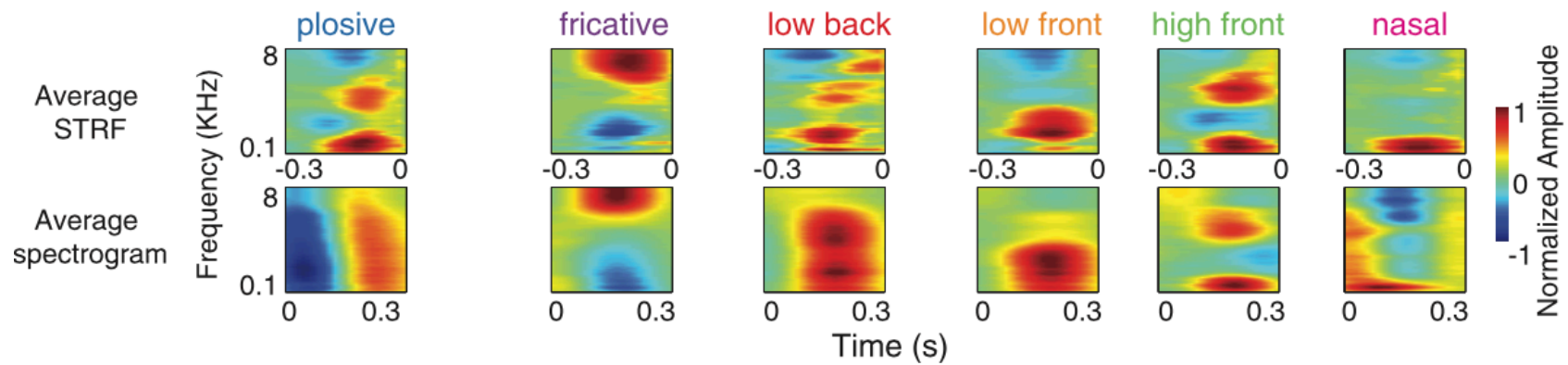
Jenison et al, 2015

STRFs as efficient coding transform

- “...STRFs result from an optimal tradeoff between maximizing the sensory information the brain receives, and minimizing the cost of the neural activities required to represent and transmit this information. Both terms depend on the statistical properties of the sensory inputs and the noise that corrupts them ...”

(Zhao & Zhaoping, 2011)

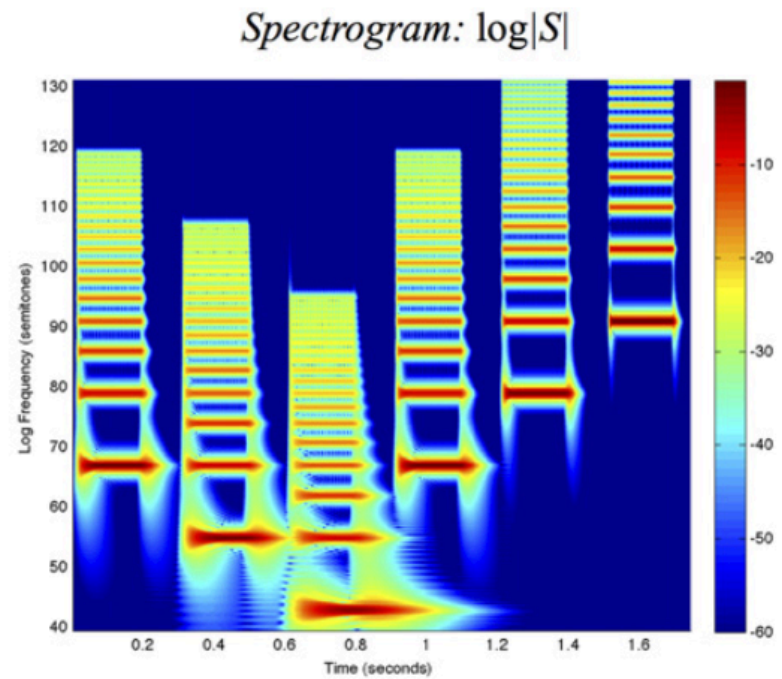
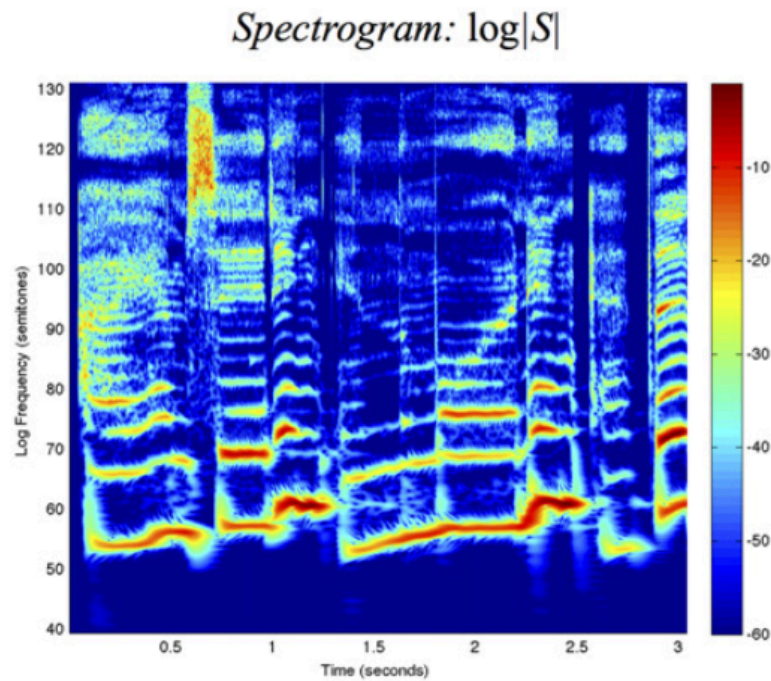
STRFs as phonetic feature detectors



Mesgarani et al, 2014

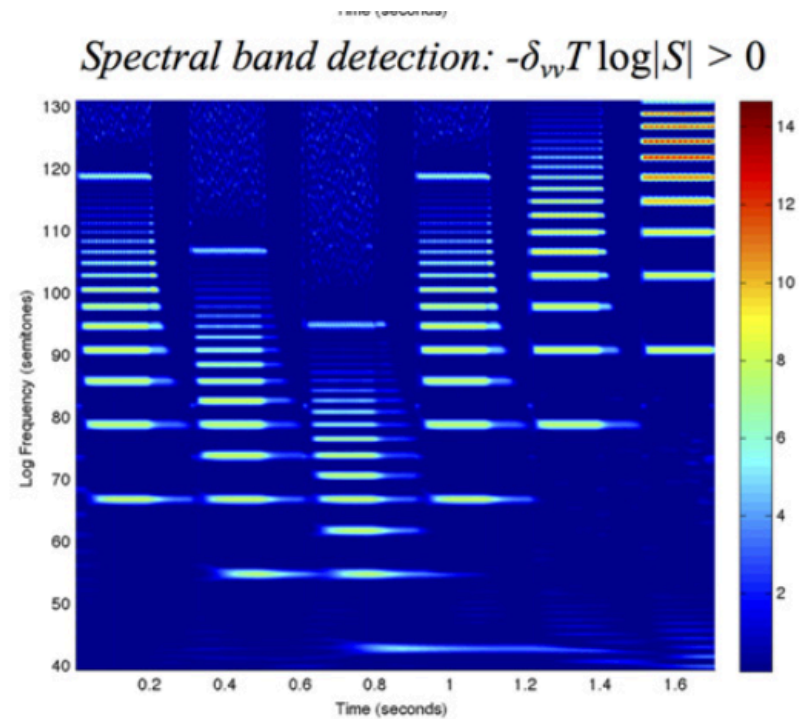
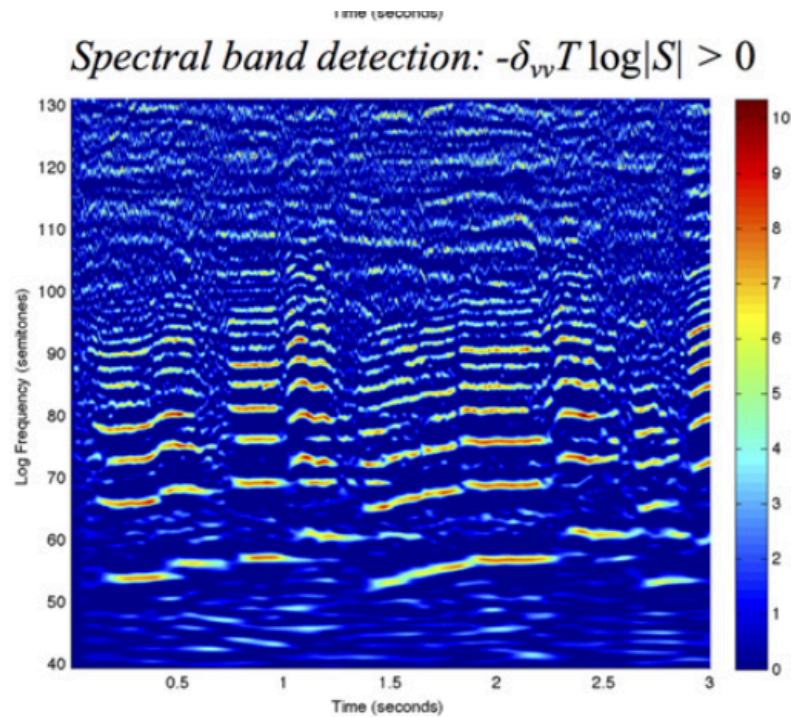
linear operators

STRFs as linear operators

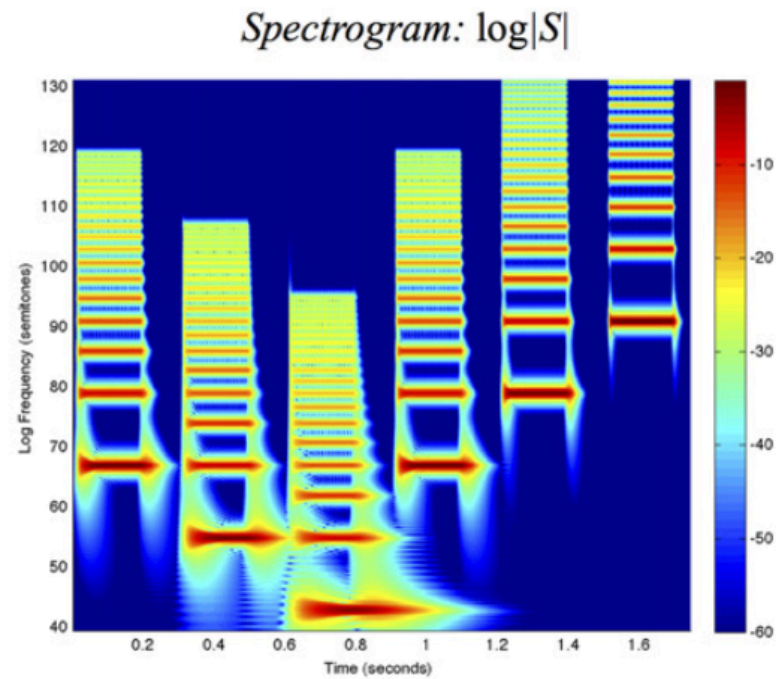
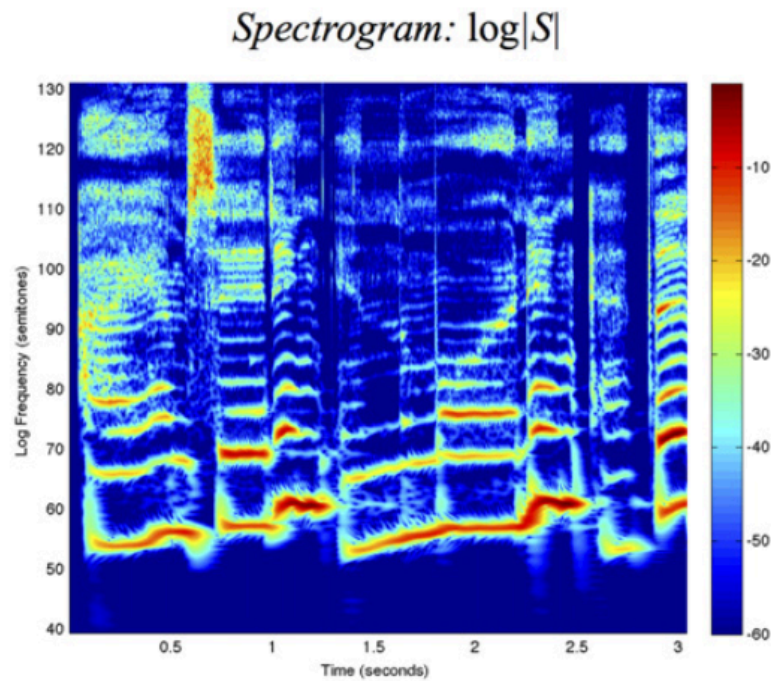


Lindeberg & Friberg, 2015

STRFs as linear operators

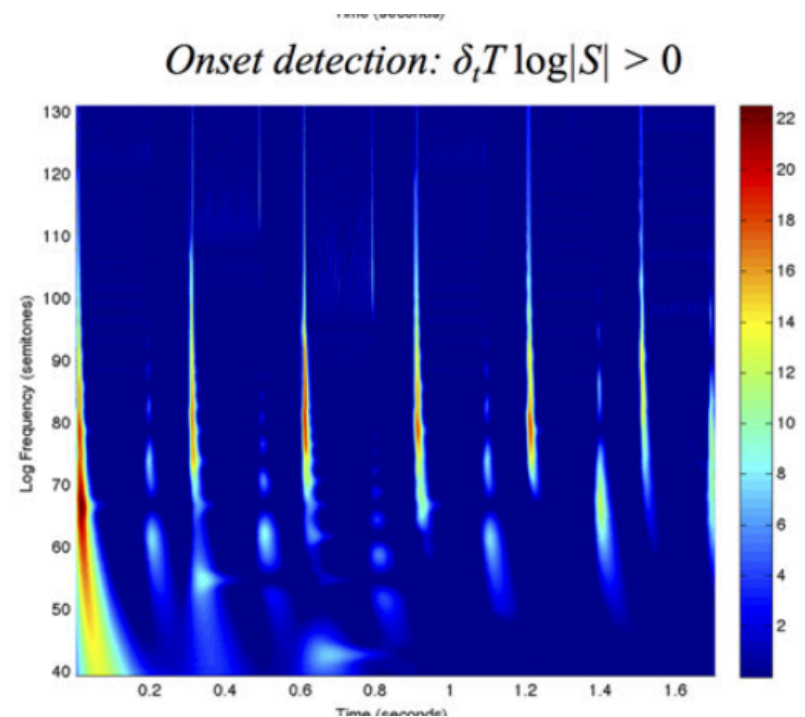
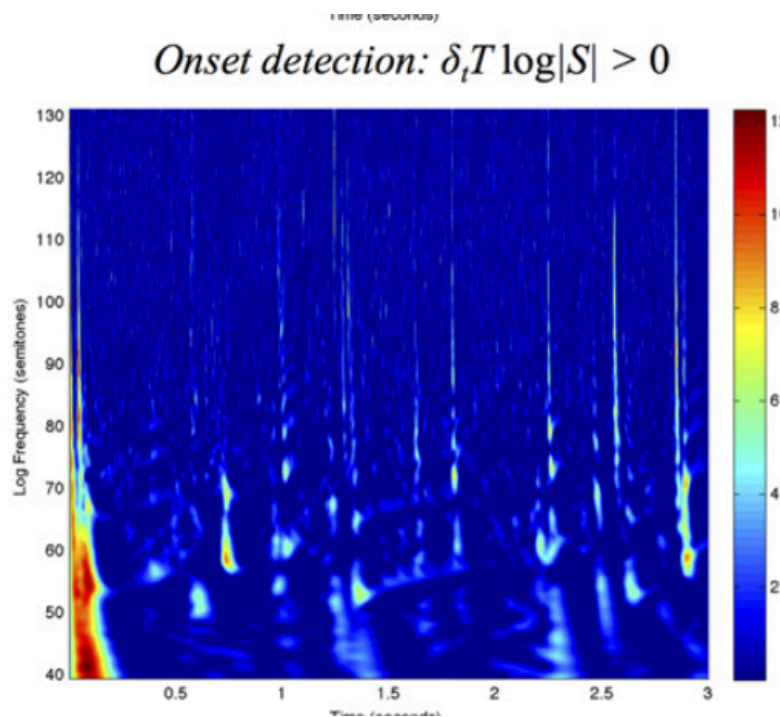


STRFs as linear operators



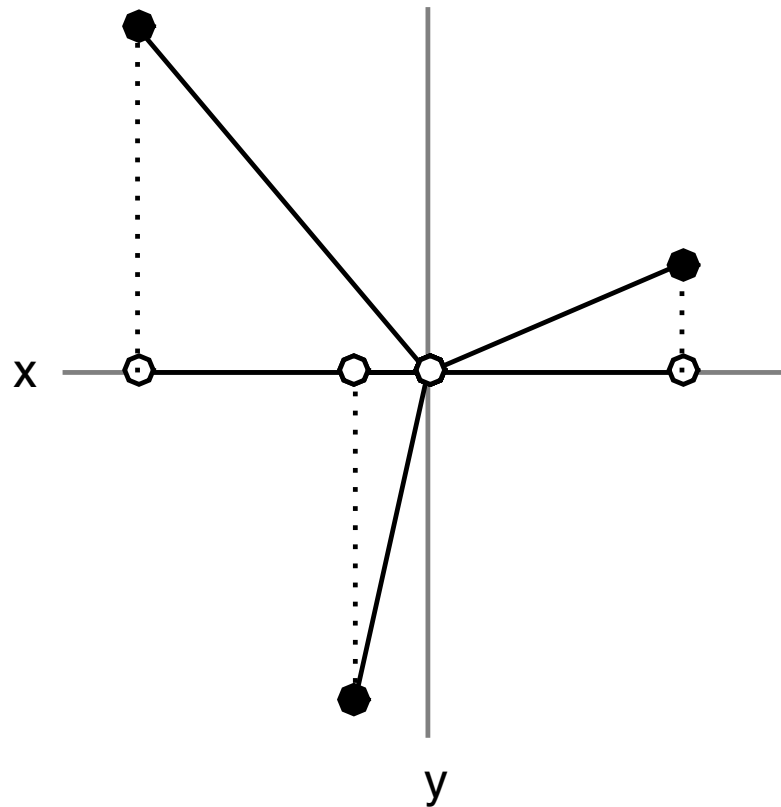
Lindeberg & Friberg, 2015

STRFs as linear operators



Lindeberg & Friberg, 2015

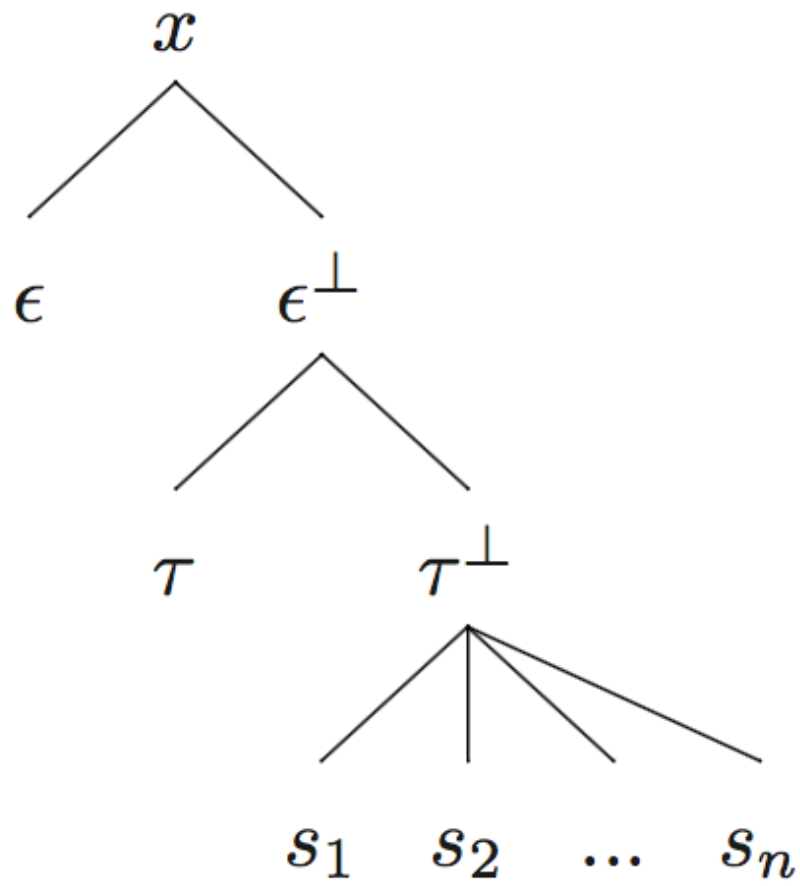
Projection operators



Working hypothesis

- Phonetic categories are vector subspaces
- Vectors are the complex-valued bark-scaled spectra
- Each subspace (category) comes with a projection operator
- Phonetic identity is recovered by applying projection operators to input signals and measuring the extent to which they project onto that subspace

model



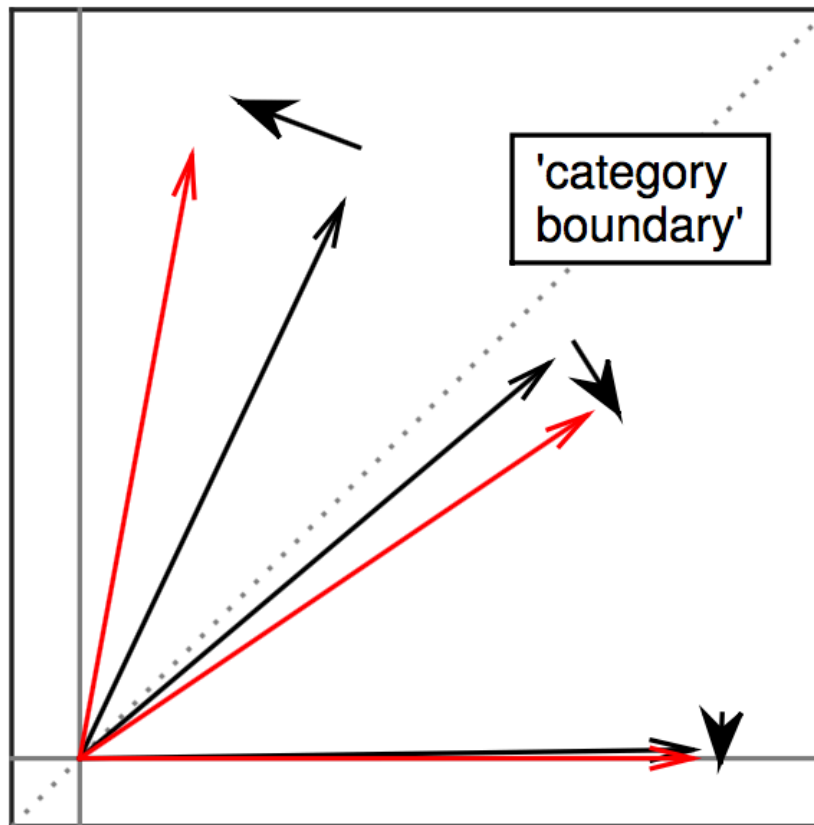
equations

$$w_i = \left\langle \frac{s_i}{\|s_i\|}, \frac{\tau^\perp}{\|\tau^\perp\|} \right\rangle^g = \frac{\|s_i\|^g}{\|\tau^\perp\|^g}$$

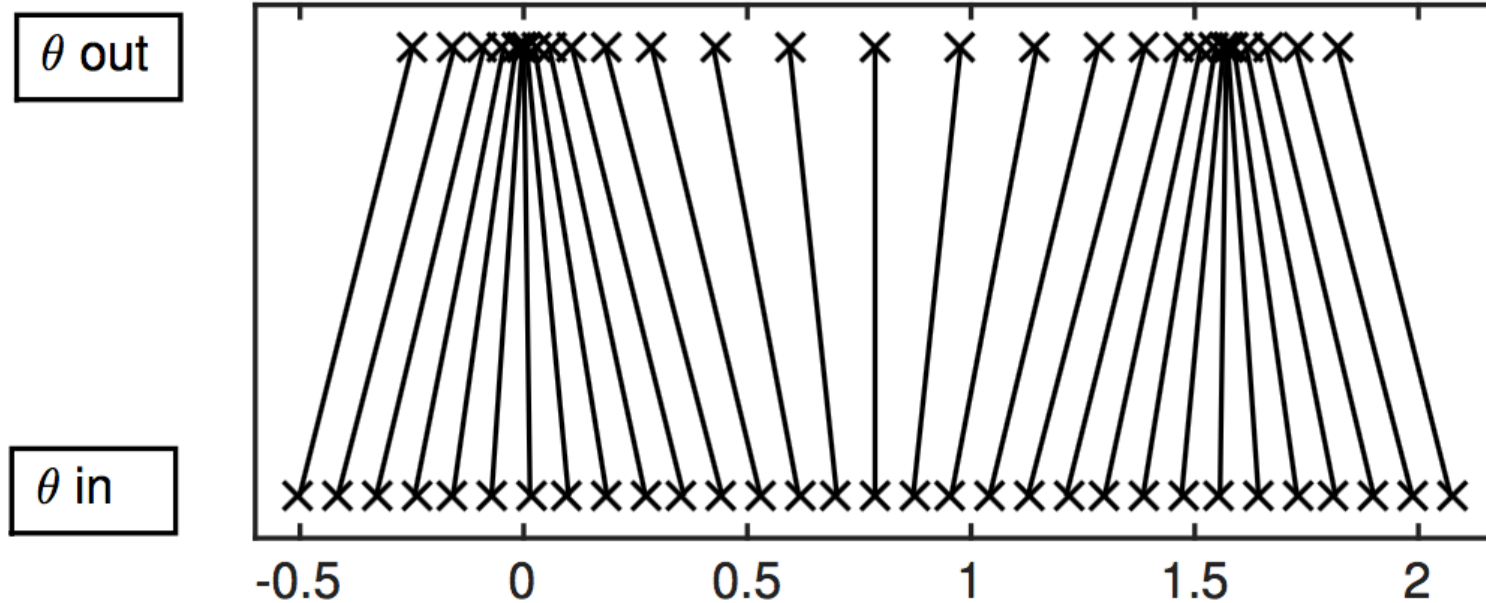
$$p_i = \frac{w_i}{\sum_{j=1}^N w_j}$$

$$\hat{x} = \sum_{i=1}^N p_i s_i$$

in 2D with standard bases



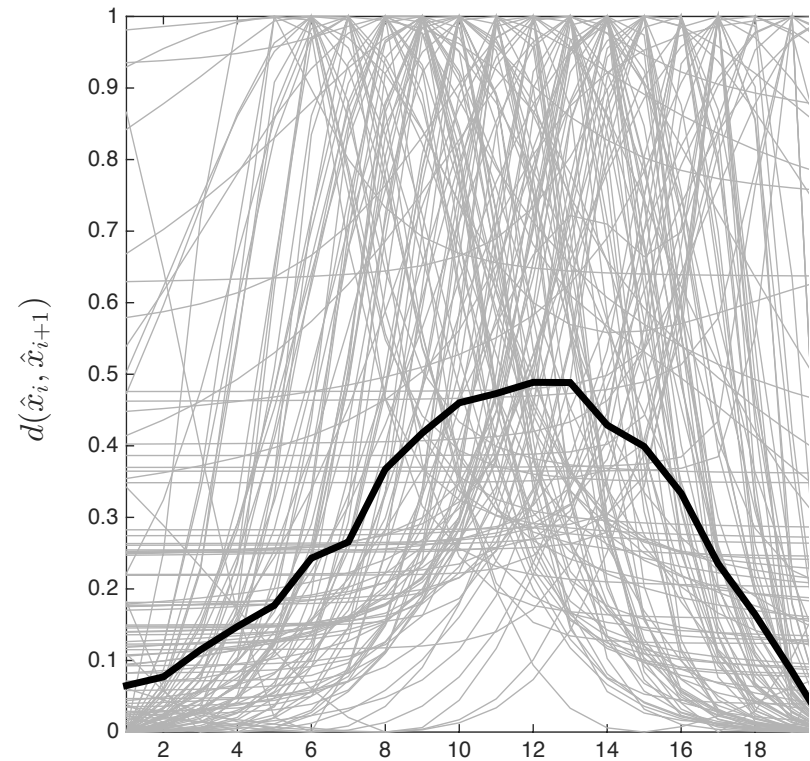
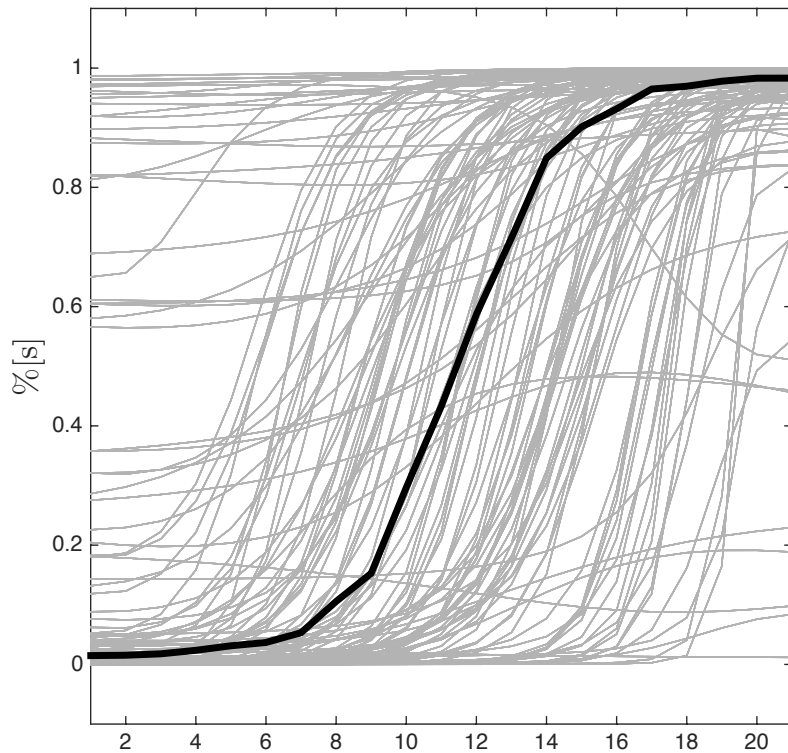
in 2D with standard bases



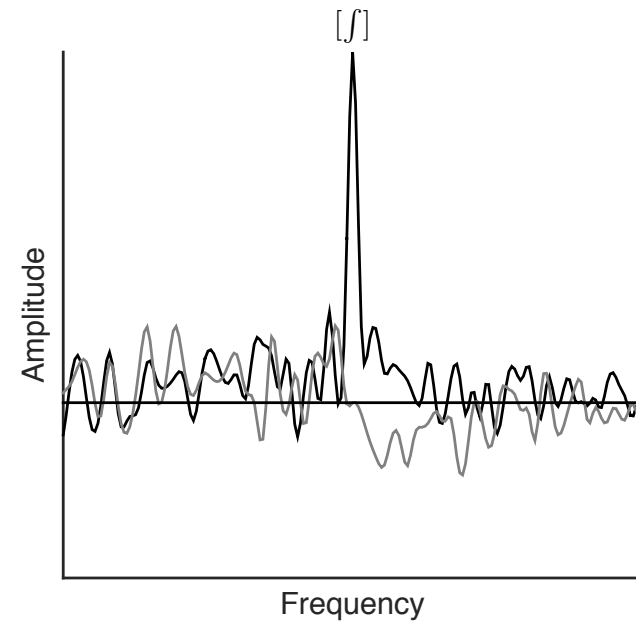
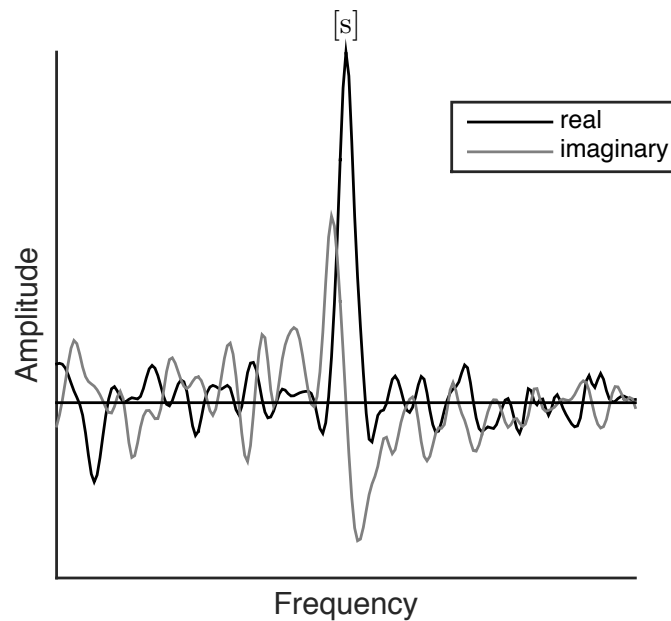
simulation

- Every [s] and [sh] in TIMIT training dataset used to estimate projection operators
- For every speaker in the test data set with at least 1 [s] and [sh], a linear continuum constructed, and probabilities and distances between successive items in the continua computed

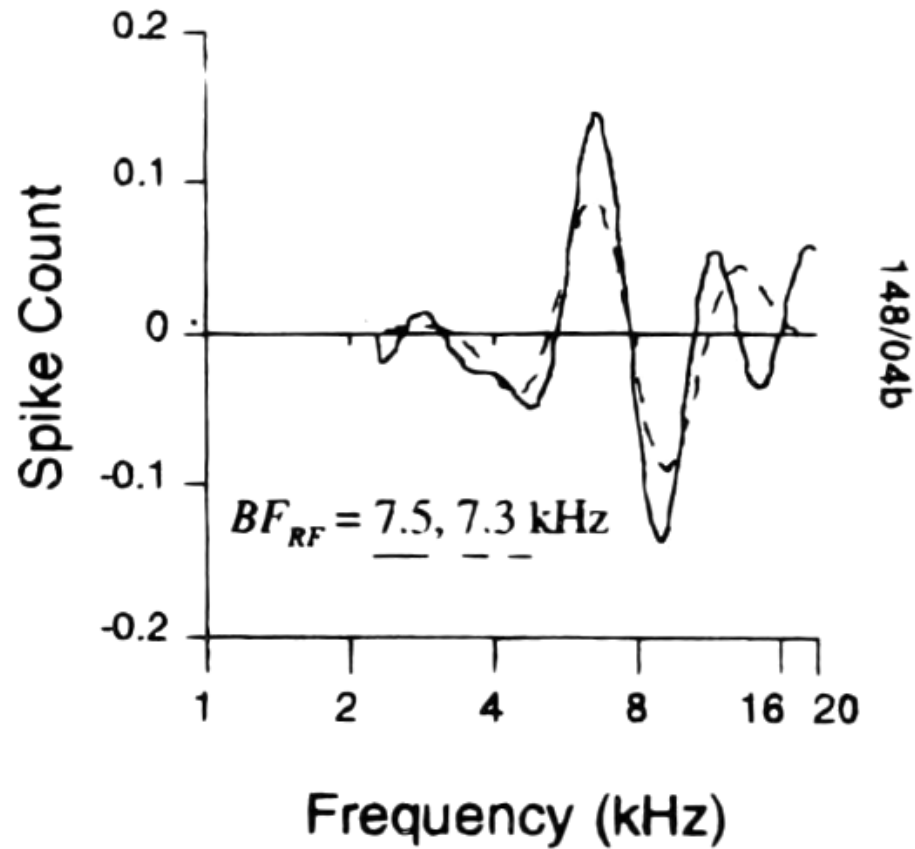
ID and discrimination functions



SRFs from projection operators



A1 (ferret)



Shamma et al, 1995

Acknowledgements

- Bill Idsardi, Naomi Feldman
- NSERC

References

- Andoni, S., Lin, N., Pollack, G.D. (2007). Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *Journal of Neuroscience* 27: 4882-4893.
- Atencio, C.A. and Schreiner, C.E. (2012). Spectrotemporal processing in spectral tuning modules of cat primary auditory cortex. *Plos ONE*, 7:e31537.
- Jenison, R.L., Reale, R.A., Armstrong, A.L., Oya, H., Hiroto, K., Howard, M.A.H.III. (2015). Sparse Spectro-Temporal Receptive Fields Based on multi-unit and high-gamma responses in human auditory cortex. *Plos ONE* 10(9).
- Kiang, N.Y.S. (1975). Stimulus representation in the discharge pattern of auditory neurons. In *The Nervous System vol 3: Human Communication and its disorders*. Raven Press, New York.
- Liberman, A., Harris, K., Hoffman, H., and Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5):358– 368.
- Mesgarani, N., Cheung, C., Johnson, K. and Chang, E.F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343: 1006-1010.
- Lindeberg, T. and Friberg, A. (2015). Idealized Computational Models for Auditory Receptive Fields. *Plos ONE*: 10(3).
- Shamma, S.A., Versnel, H. and Kowalski, N. (1995). Ripple analysis in Ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Auditory Neuroscience*, 1: 233-254.
- Sutter, M.L., Schreiner, C.E., McLean, M., O'Connor, K.N. and Loftus, W.C. (1999). Organization of inhibitory frequency receptive fields in cat primary auditory cortex. *Journal of Neurophysiology*, 82: 2358-2371
- Wang, S., Hamilton, A., Koickal, T.J. Cheung, R. and Smith, L.S. (2014). A Bio-Realistic Analog CMOS cochlea filter with high tunability and ultra-steep roll-off. *IEEE Transactions on Biomedical Circuits and Systems* 9(3).
- Zhao, L. and Zhaoping, L. (2011). Understanding Auditory Spectro-Temporal Receptive Fields and Their Changes with Input Statistics by Efficient Coding Principles. *PLoS Computational Biology*: 7(8).