

Spectro-Temporal Methods in Primary Auditory Cortex

David Klein
Didier Depireux
Jonathan Simon
Shihab Shamma

Institute for Systems Research
& Department of Electrical Engineering
University of Maryland

Supported in part by a MURI grant from the
Office of Naval Research and by a grant from
the National Science Foundation

also available at <<http://www.isr.umd.edu/CAAR/pubs.html>>



Summary

- In Primary Auditory Cortex (AI) of ferrets, we characterize the spectro-temporal properties of cells' responses. We find that the responses correspond to temporal modulations from 4 to 40 Hz, and spectral modulations from 0 to 2 cycles/octave, in the stimulus' spectro-temporal envelope.
- The Spectro-Temporal Response Function (**STRF**) is the linear component of the response. It is an excellent predictor of response.
- Different methods of determining the STRF are in good agreement, make similar (and similarly accurate) predictions.
- Spike-Triggered Averaging is an effective method to measure the STRF, when used with Temporally Orthogonal Ripple Combinations (TORCs) as stimuli.
- Spike-Triggered Averaging methods do not depend on quadrant separability, and provide a good method for seeking non-quadrant separable responses.

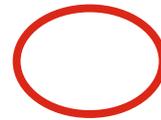
Auditory Stages and Representations

Cochlea



- Sound is bandpassed, half-wave rectified, and low-passed

Inferior Colliculus



- Envelope of the sound is bandpassed or lowpassed, extracting the first (fast) modulation of the sound near BF

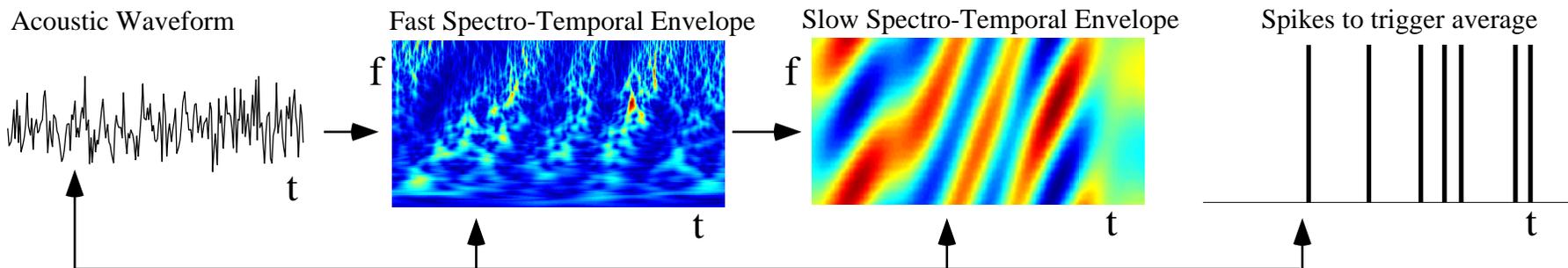
Cortex



- Band-passed envelope from IC is low-passed, extracting the second (slow) modulation of the sound near BF

Each stage of the auditory pathway represents the auditory stimulus differently. After cochlear processing, the acoustic waveform is represented by parallel, frequency-ordered, neural signals, which can encode different characteristics of the sound in different areas. By primary auditory cortex, responses represent the slow modulations (a few to a few tens of Hertz) of the envelope of the signal.

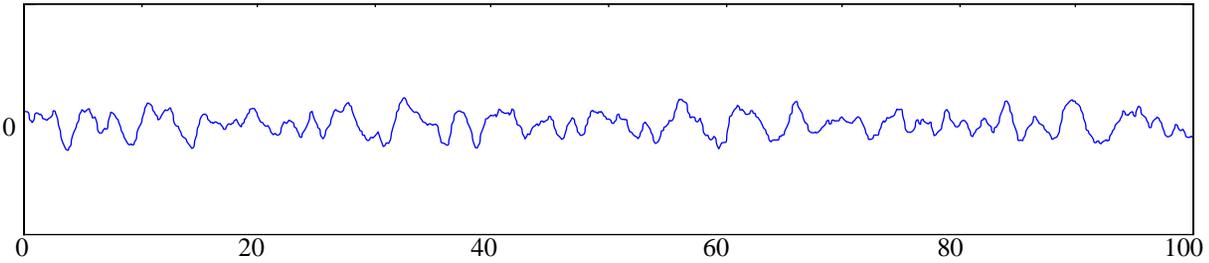
Spike-triggered averaging, a method for extracting the linear component of the response, can be done on neural signals at any stage, and the averaging can be on any representation of the signal. The stimuli must be “rich” for the final average to fully characterize the response.



Introduction to Cortical Responses I

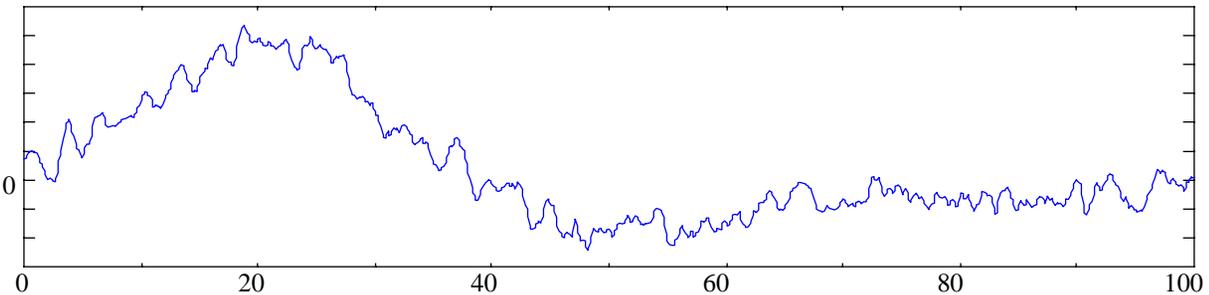
Spike-triggered average waveform

Inability of cortical cells to lock to waveform usually renders this statistic useless



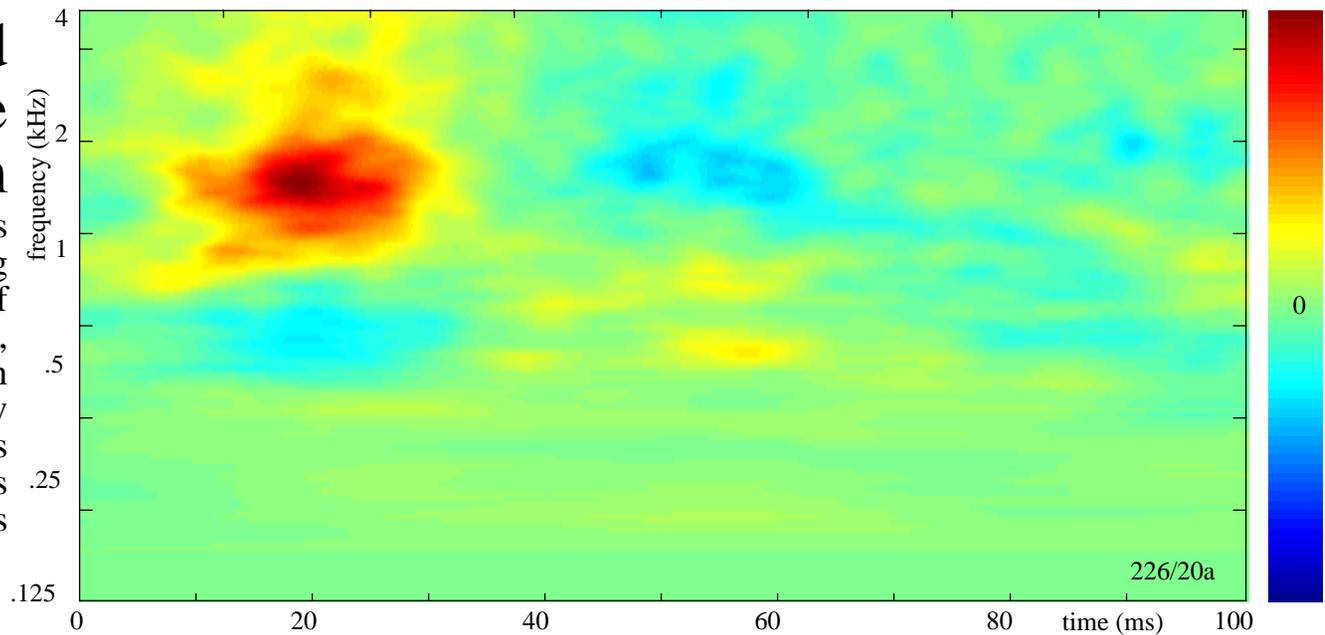
Spike-triggered average envelope

Cortical cells lock to the slow(er) modulations of the envelope. The average does not give much frequency selectivity information



Spike-triggered average spectrogram

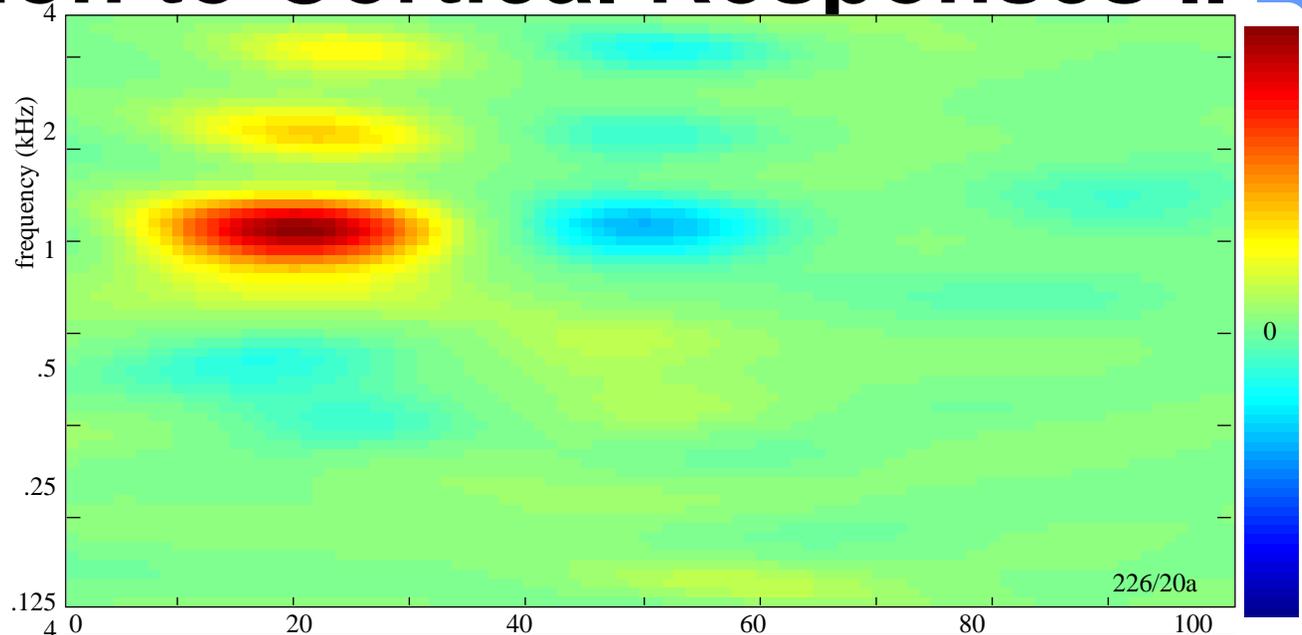
Bandpassing at various frequencies before computing the spike-triggered average of the envelope gives more detail, such as differentiating between inhibitory and excitatory frequency bands, and shows different temporal responses for different frequency bands



Introduction to Cortical Responses II

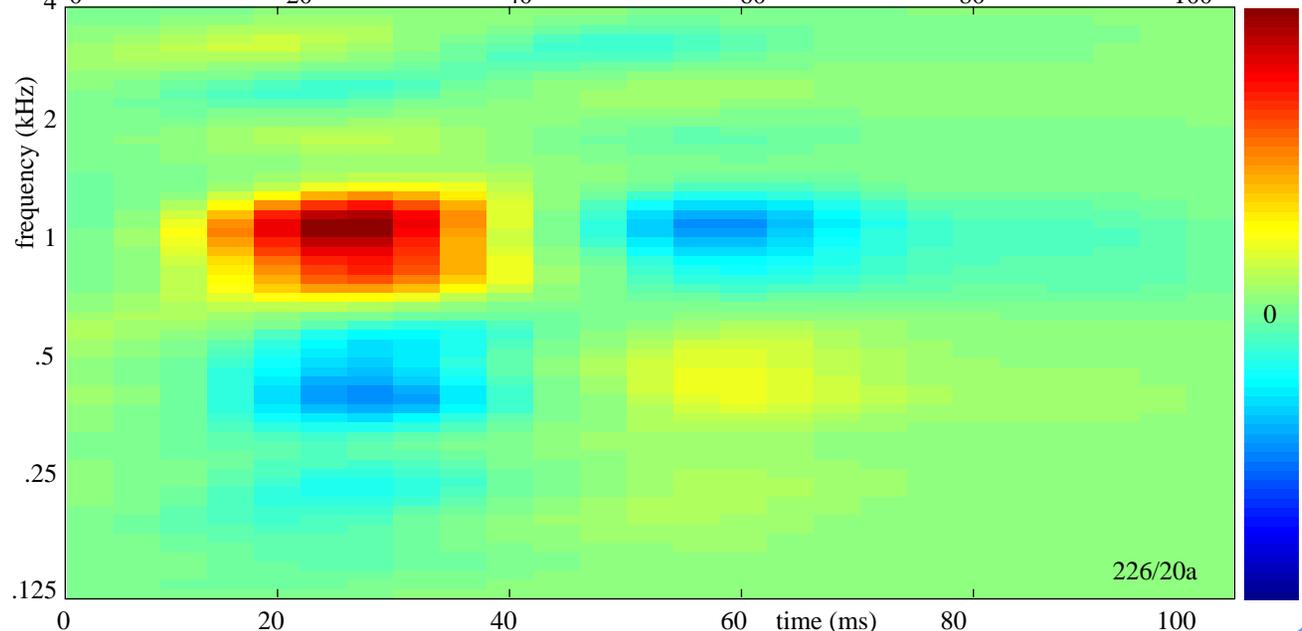
Spike-triggered spectro-temporal envelope average

Spike-triggered averaging of the spectro-temporal envelope directly gives a similar spectro-temporal response field to the spike-triggered average of the filter-bank envelopes



STRF from Ripple Transfer Function

The spectro-temporal transfer function is compiled by measuring the response amplitude and phase to single sinusoidally modulated spectra that move in time (ripples). The spectro-temporal response function is the 2-D inverse Fourier transform.

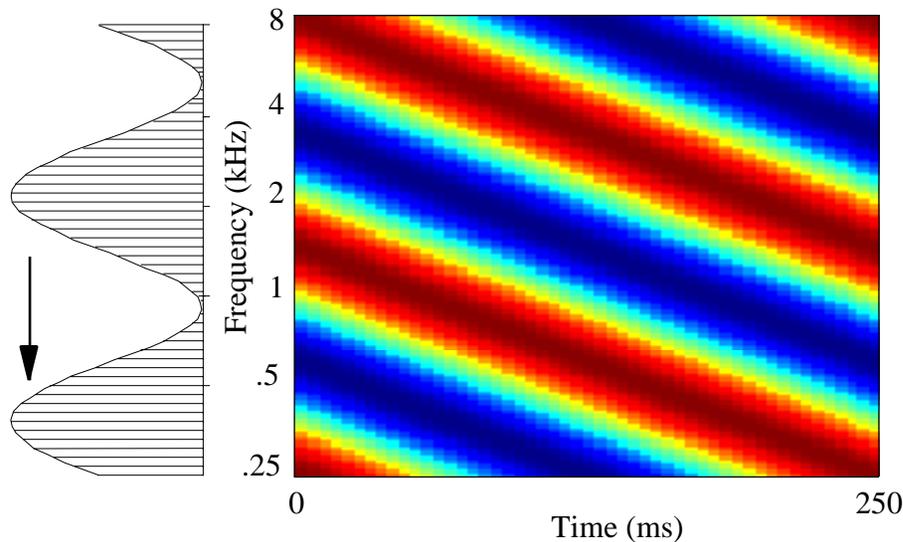


Motivation and Methods

- Primary cortical cells prefer spectral envelope modulations from 4 to 40 Hz, and 0 to 2 cycles/octave.
- **Ripple transfer functions**, the Fourier transform of the **Spectro-Temporal Response Field (STRF)**, are time-consuming to measure. Assuming **separability** (or at least quadrant-separability), measurement time is reduced, but how universal is separability?
- **STRF** can also be measured by **spike-triggered averaging** of spectro-temporal representations of the stimulus. The stimulus must be “rich” in the spectro-temporal modulations that characterize the response.
- **Spike-triggered averaging** does not rely on separability and can be used to test separability directly.
- As stimuli, Temporally Orthogonal Ripple Combinations (**TORCs**) cover large regions of spectro-temporal modulation space efficiently.
 - Each stimulus is composed of superposition of moving sinusoids (ripples)
 - Stimuli have only one spectral modulation for each temporal modulation, but many temporal modulations, each a multiple of the base.
 - Stimulus components are orthogonal over the averaging interval.
 - One-to-one correspondence between stimulus components and spectral modulations removes ambiguity of which component evokes which aspect of the response dynamics.
 - Duplicating each stimulus with opposite polarity (overall sign) strongly reduces half-wave rectification non-linearity (actually all even-order non-linearities).

Spectro-Temporal Modulations

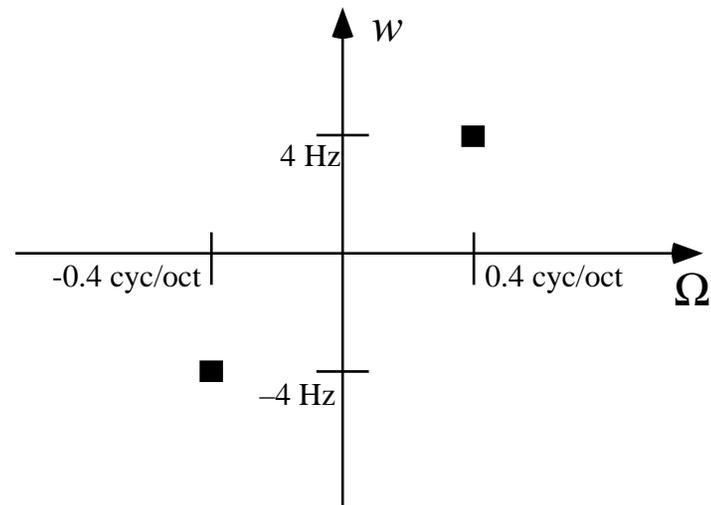
Moving Ripple in Spectro-Temporal Space (Spectrogram)



The Fourier transform of a moving sinusoid has support only on a single point (and its complex conjugate).

$$\int [.] \exp(\pm 2\pi j \Omega x \pm 2\pi j \omega t)$$

Moving Ripple in Fourier Space

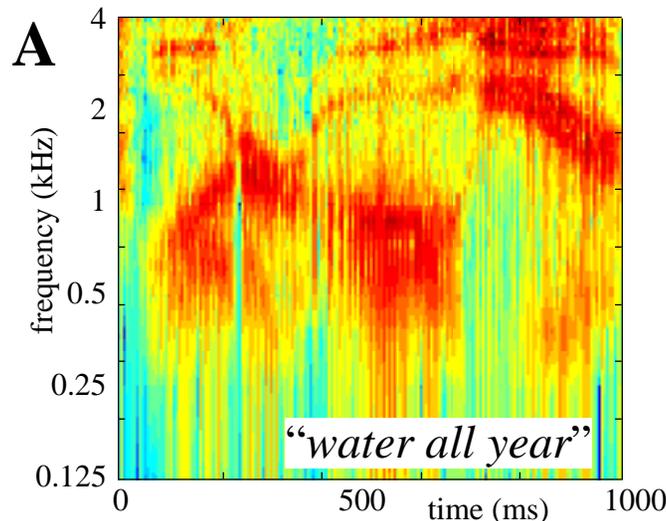


2-D Decomposition of Broadband Sound

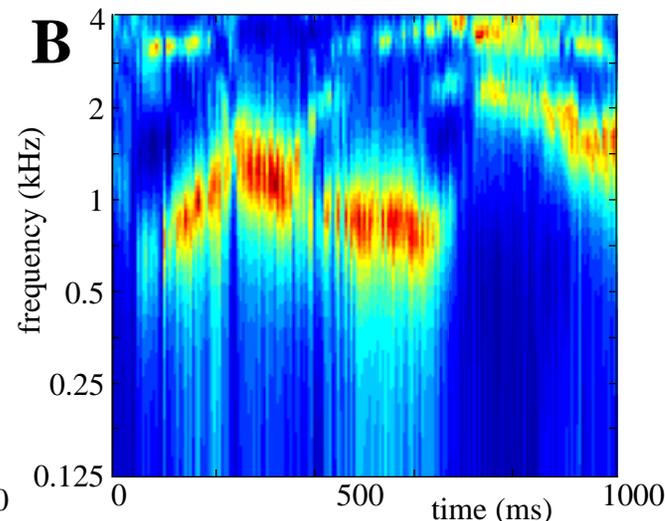
- Frequencies mapped along cochlea on log frequency axis
- Natural sounds dynamic, time axis required.
- Use two-dimensional functions of $\log(\text{freq})$ and time
- Analysis is often conceptually simpler in the Fourier domain.

(A) A speech fragment has its envelope (B) Fourier transformed. The Fourier transform is then approximated by its 100 largest components in (C) and then inverted back in (D), giving an excellent approximation to the original envelope.

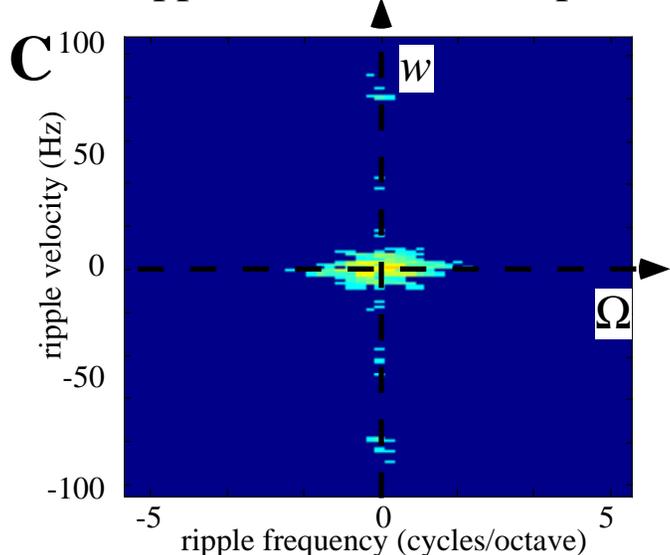
Spectrogram (log frequency)



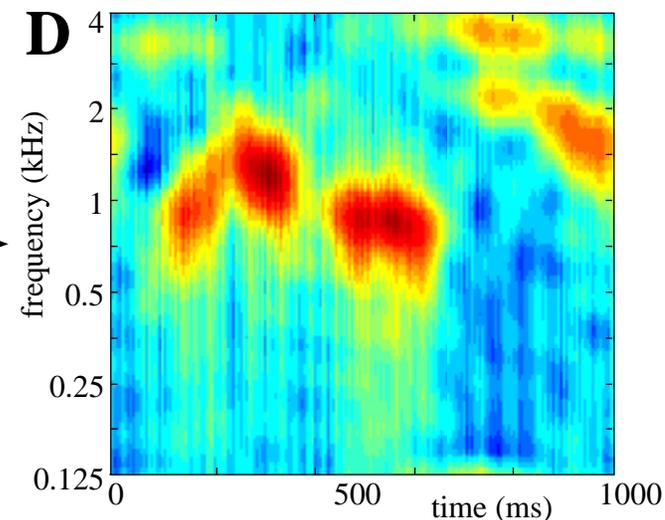
LPC Envelope



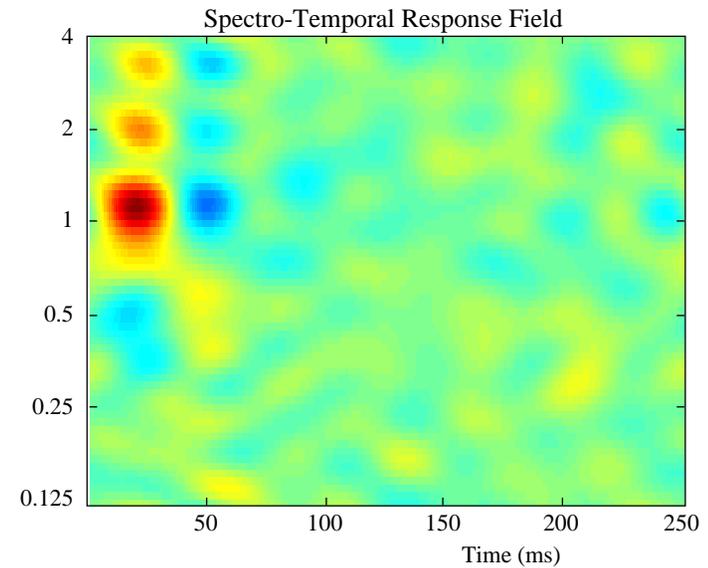
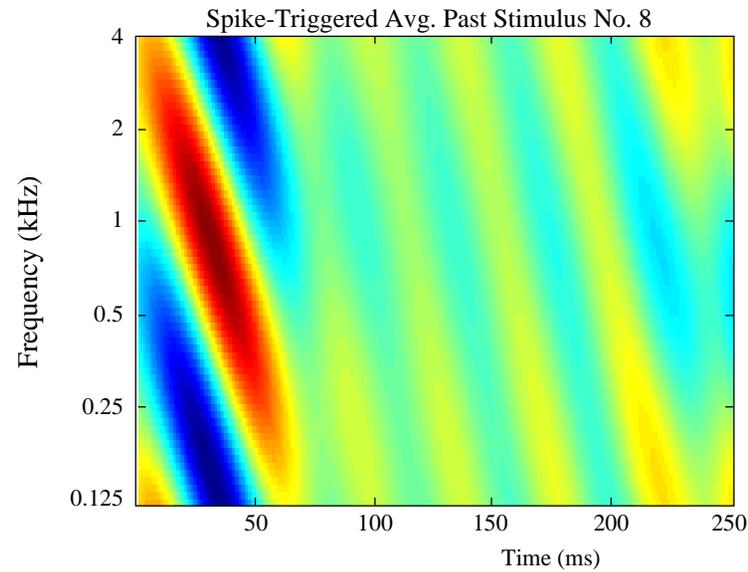
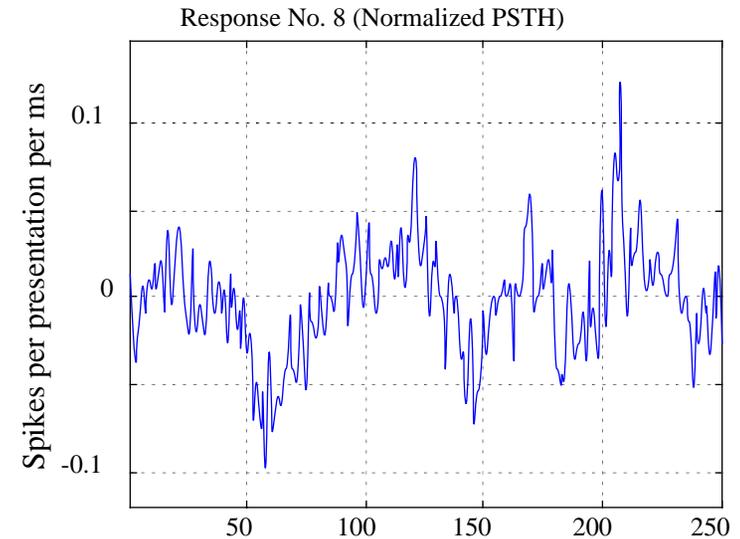
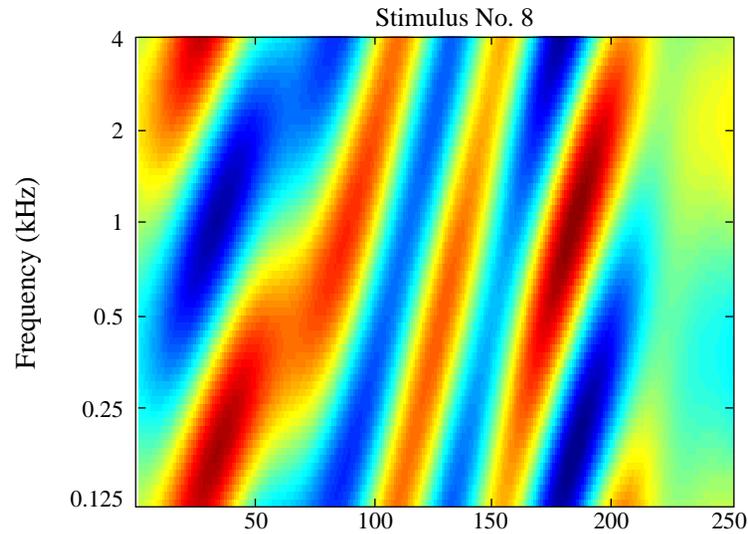
Ripple Transform (100 peaks)



Reconstruction

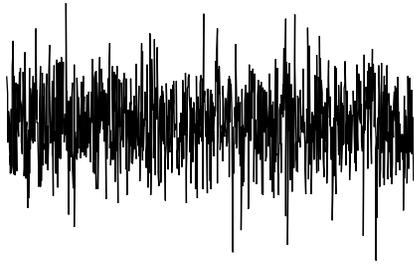


Spike Averaged Spectro-Temporal Envelope

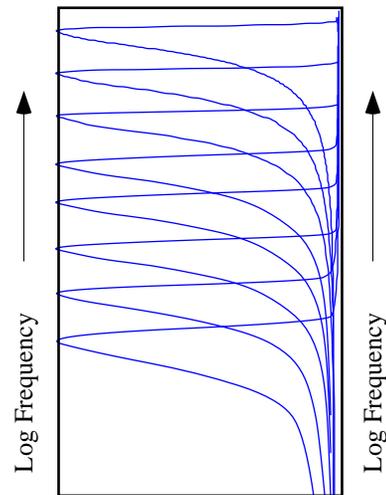


Cochlear Spectrogram

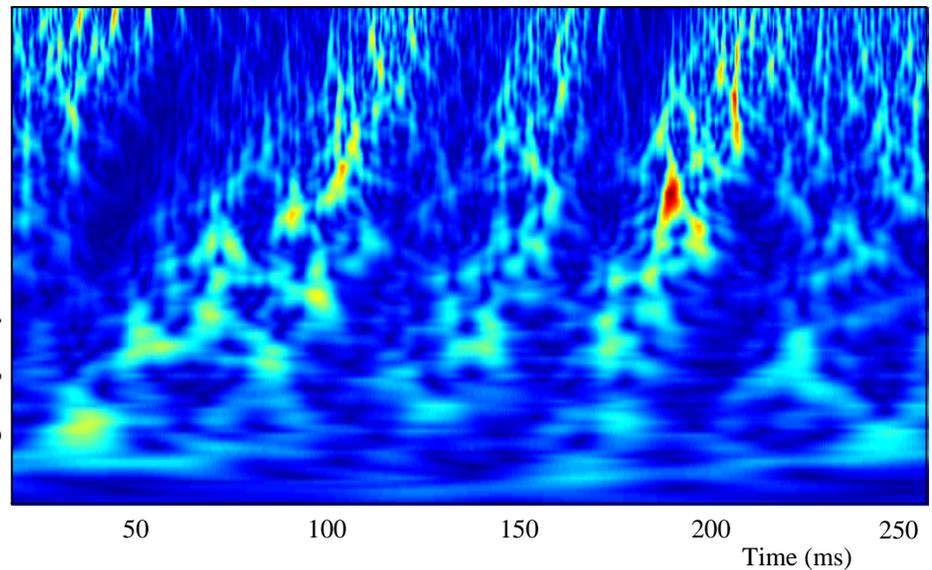
Stimulus Waveform



Cochlear Filters

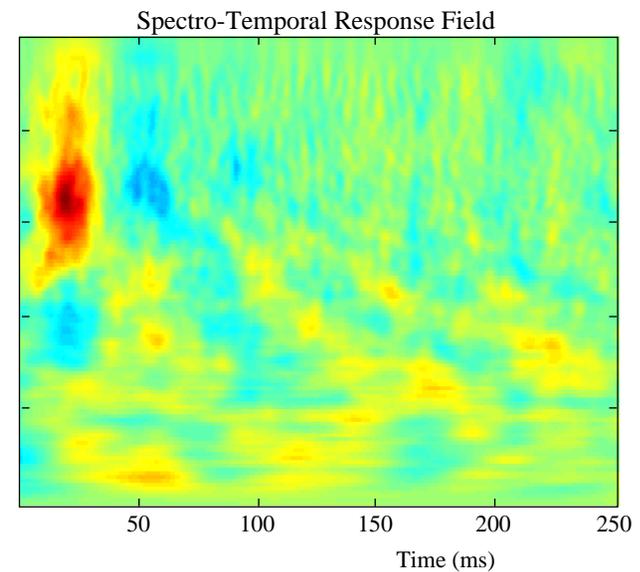
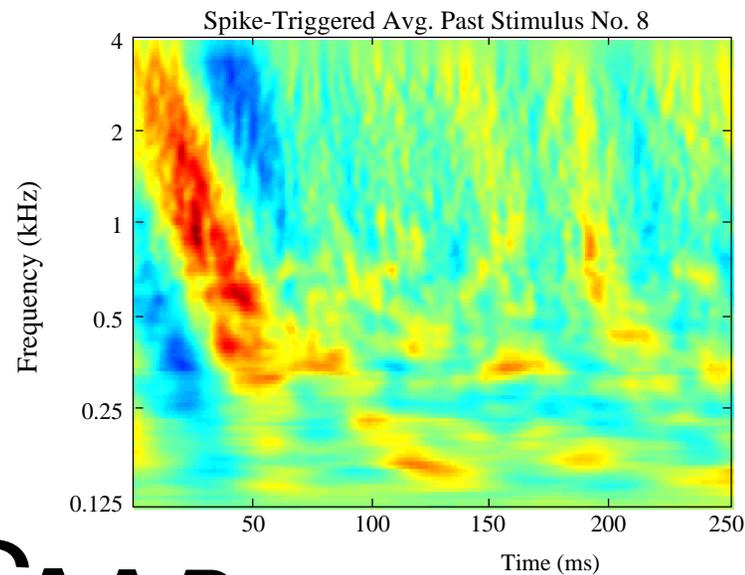
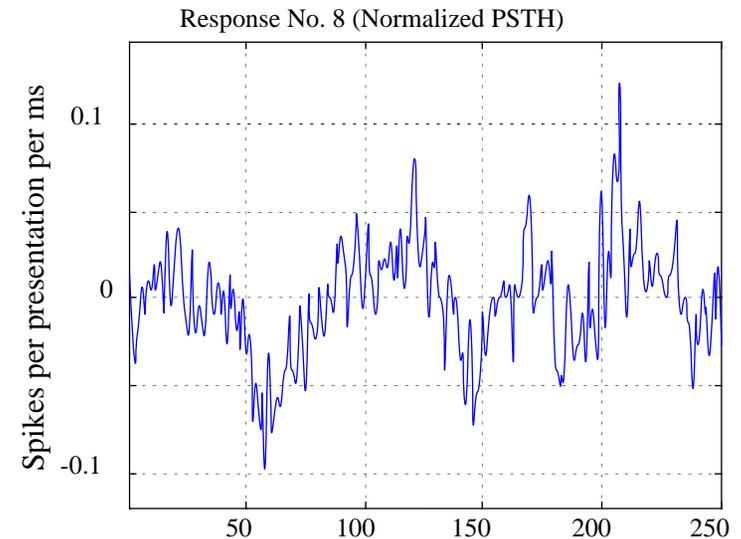
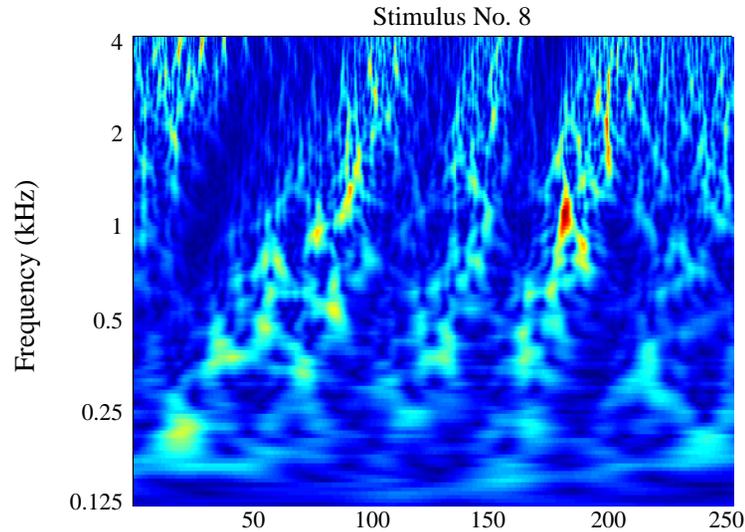


Spectrogram

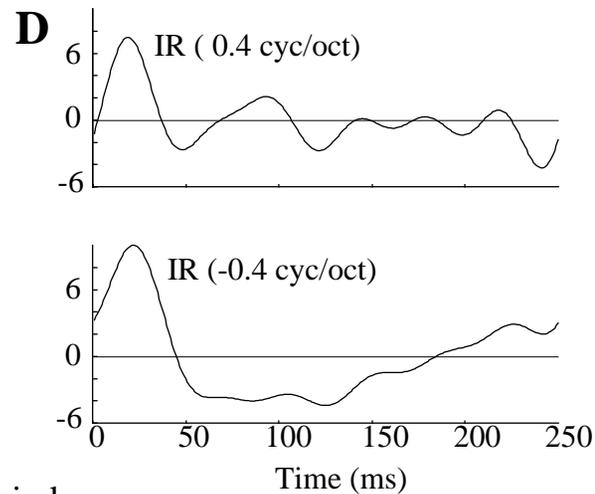
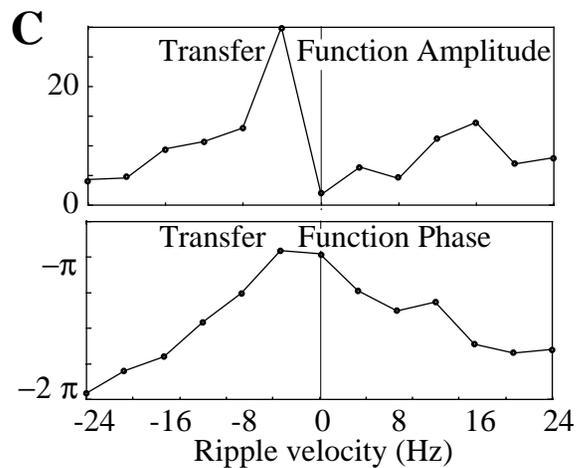
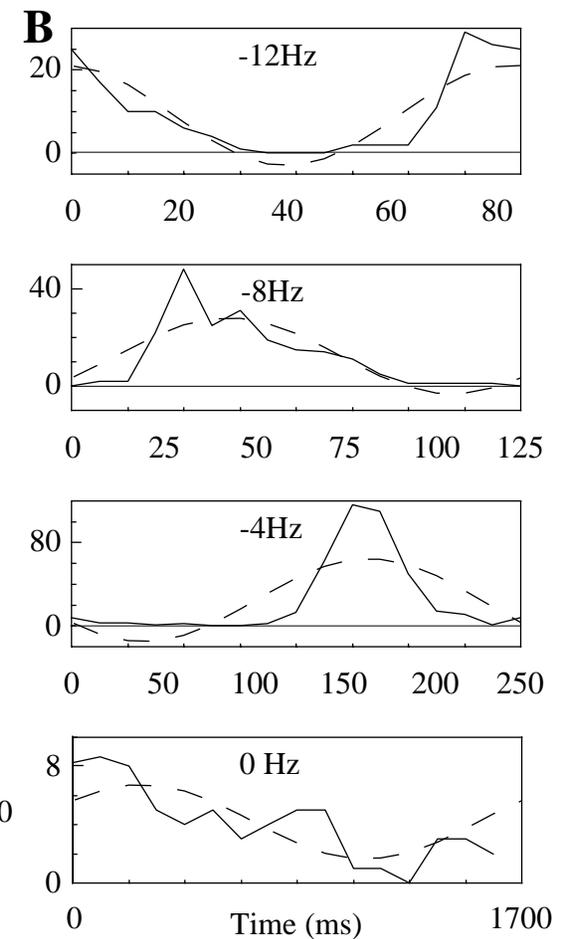
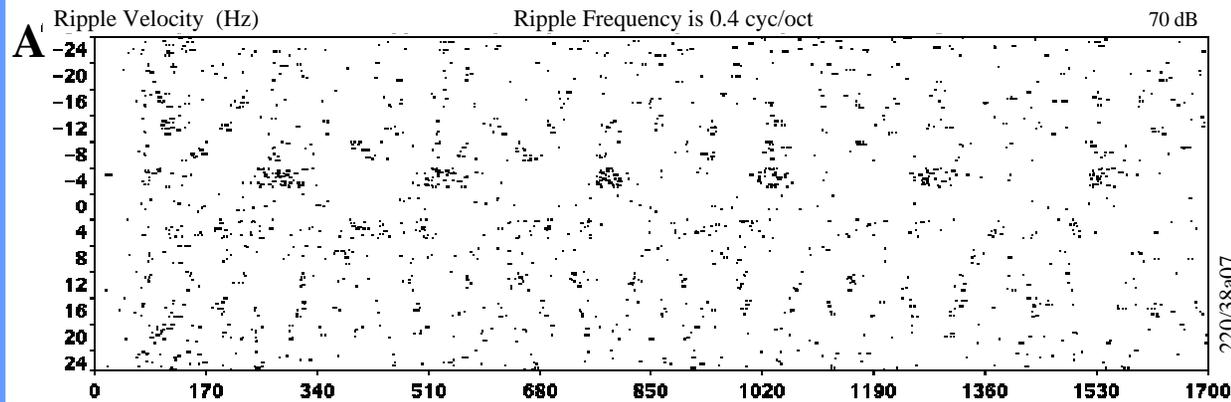


The cochlear spectrogram of the stimulus is obtained by passing the stimulus waveform through a bank of cochlear filters. The temporal envelope of each of the filter outputs form the rows of the cochlear spectrogram.

Spike-Triggered Average Spectrogram



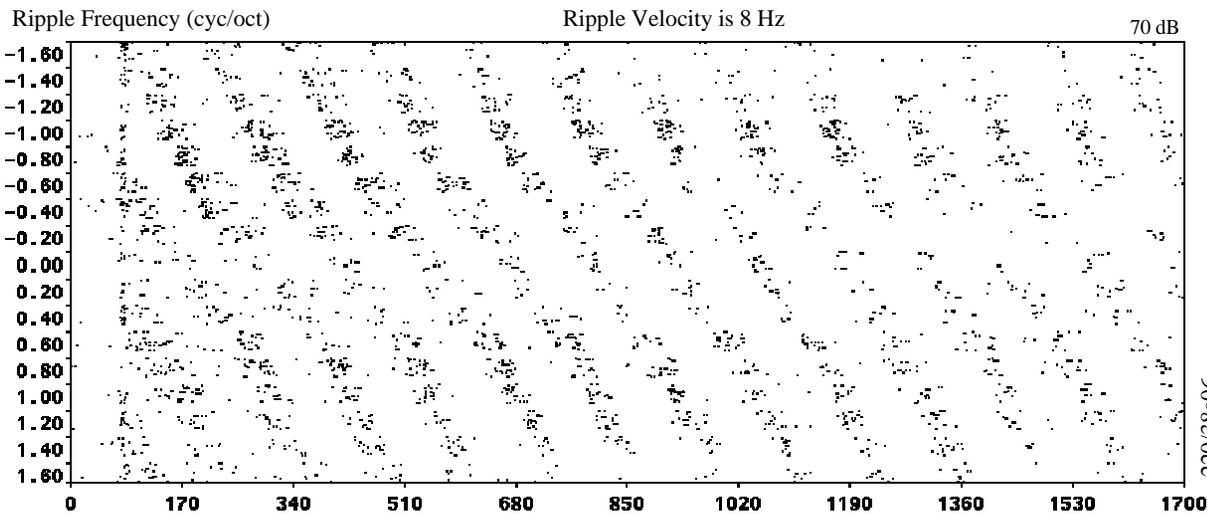
Measurements by Ripple Velocity



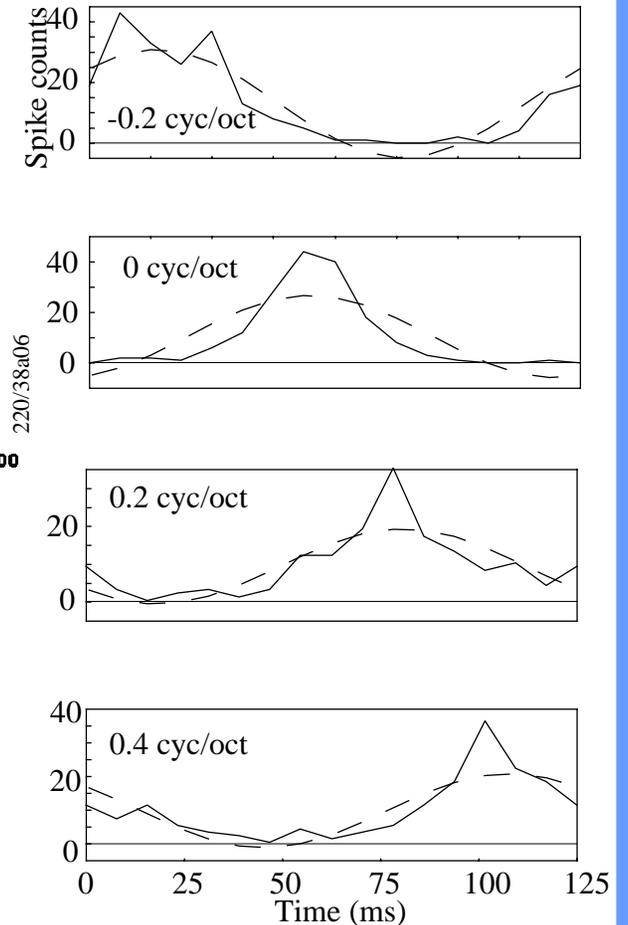
Spike events in (A) are turned into period histograms in (B). The amplitudes and phases give the transfer function in (C), which can be inverse Fourier transformed to give Impulse Responses in (D).

Measurements by Ripple Frequency

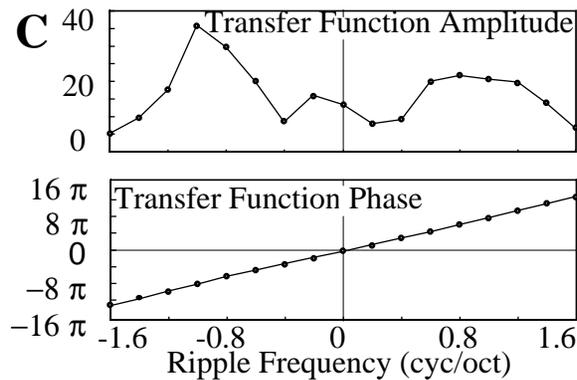
A



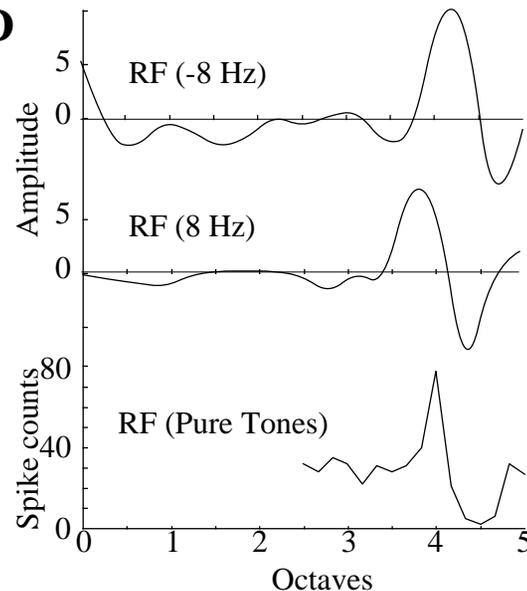
B



C



D

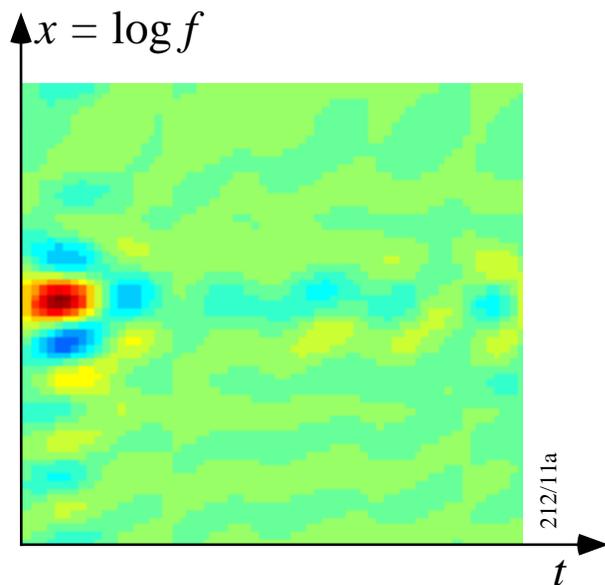


Spike events in (A) are turned into period histograms in (B). The amplitudes and phases give the transfer function in (C), which can be inverse Fourier transformed to give Response Fields in (D).

Spectro-Temporal Transfer Function

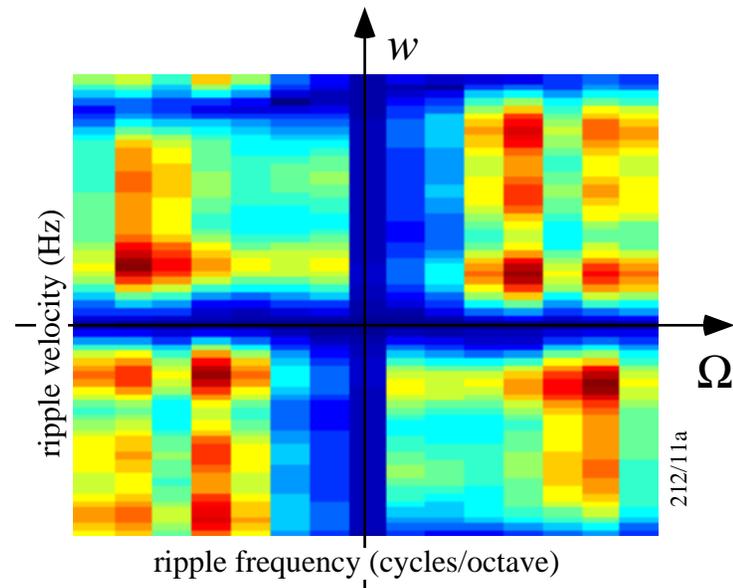
- Spectro-temporal response field of neuron is the usual response field made time-dependent.
- Its Fourier transform is the transfer function.
- Either can be used to predict the response to *any* broadband dynamic sound.

Spectro-Temporal Response Function (STRF) of a neuron

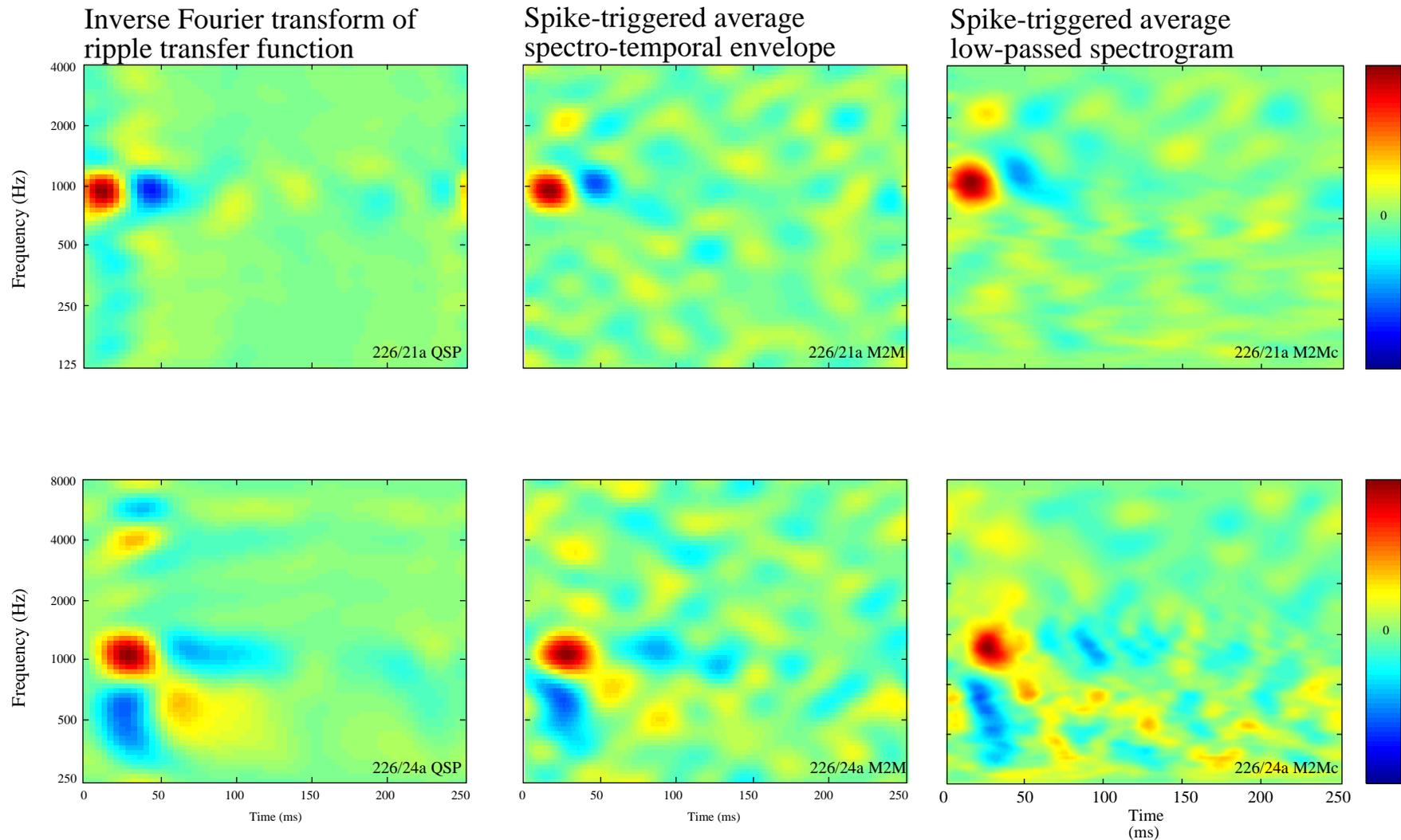


Fourier Transform
 $\int [.] \exp(\pm 2\pi j \Omega x \pm 2\pi j \omega t)$
Inverse Transform

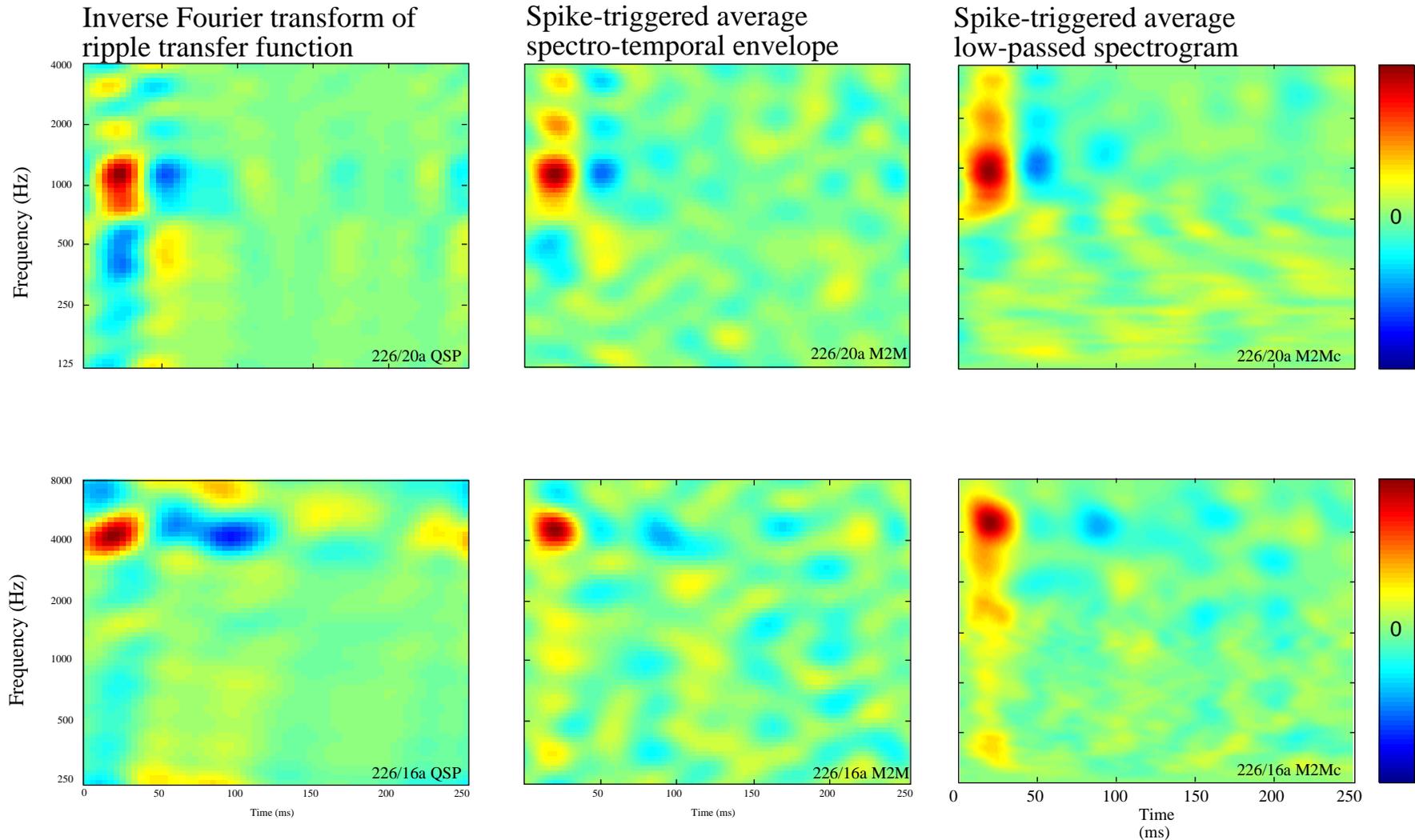
2 Dimensional Transfer Function of the same neuron



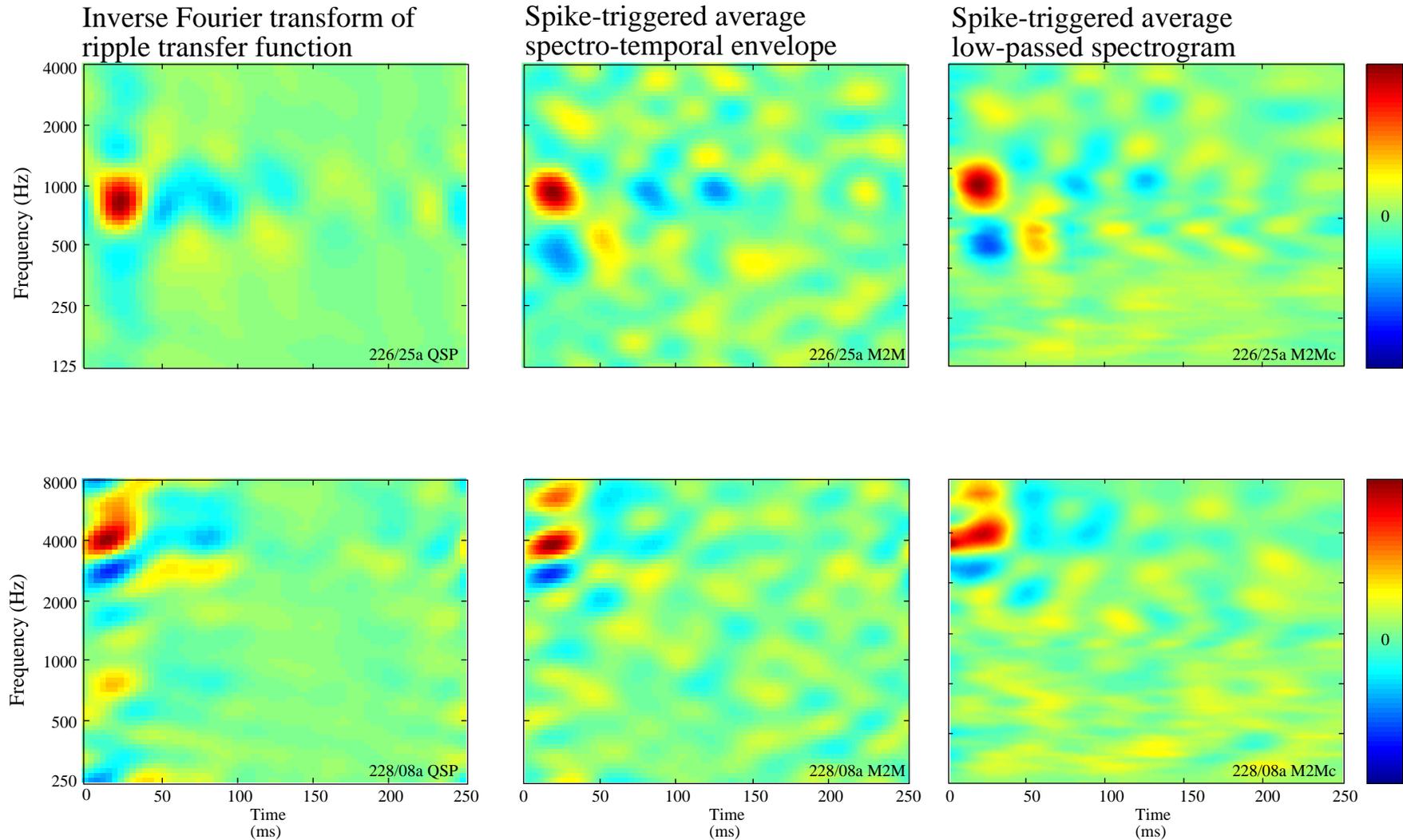
Spectro-Temporal Responses Compared I



Spectro-Temporal Responses Compared II

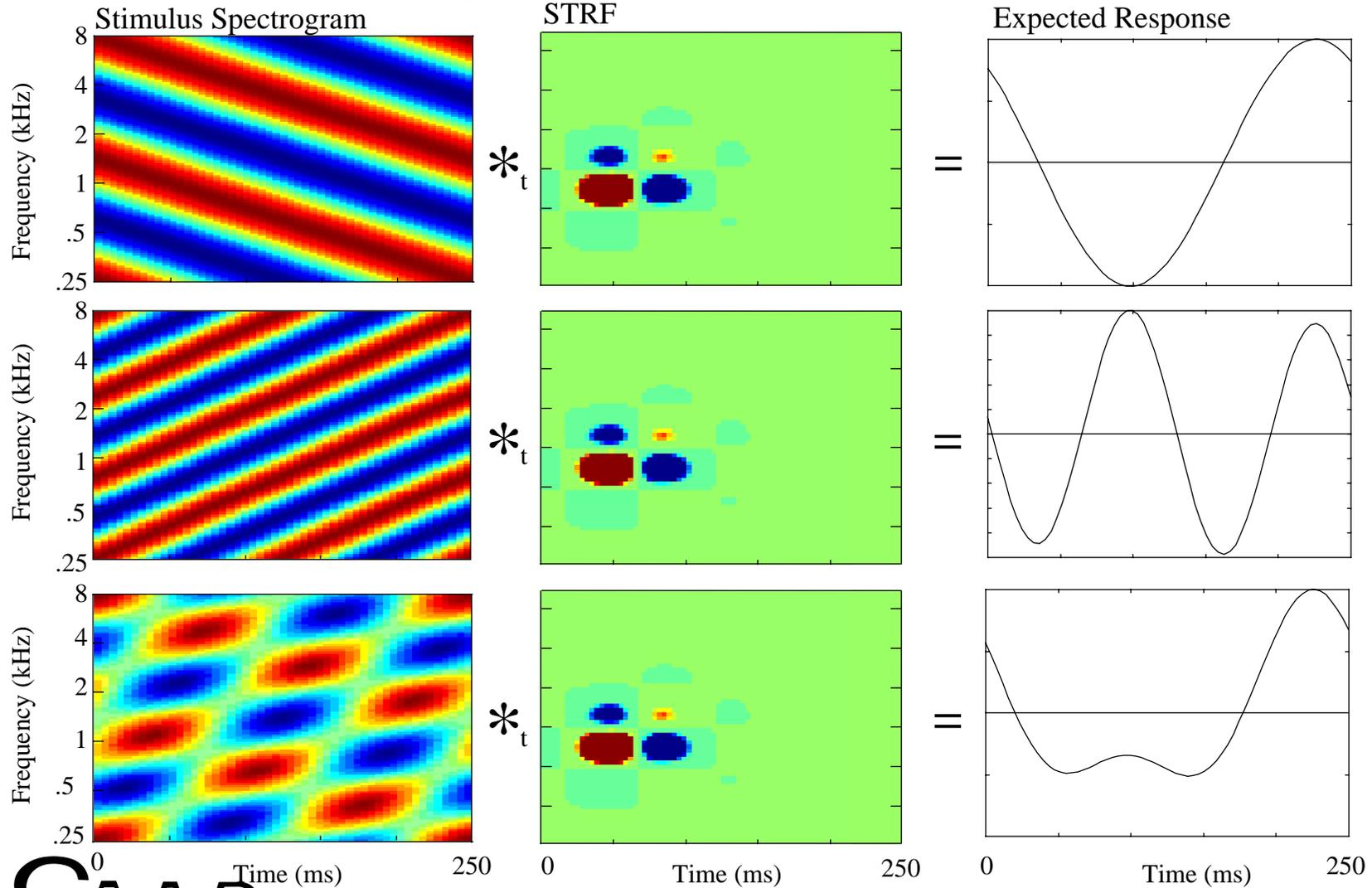


Spectro-Temporal Responses Compared III



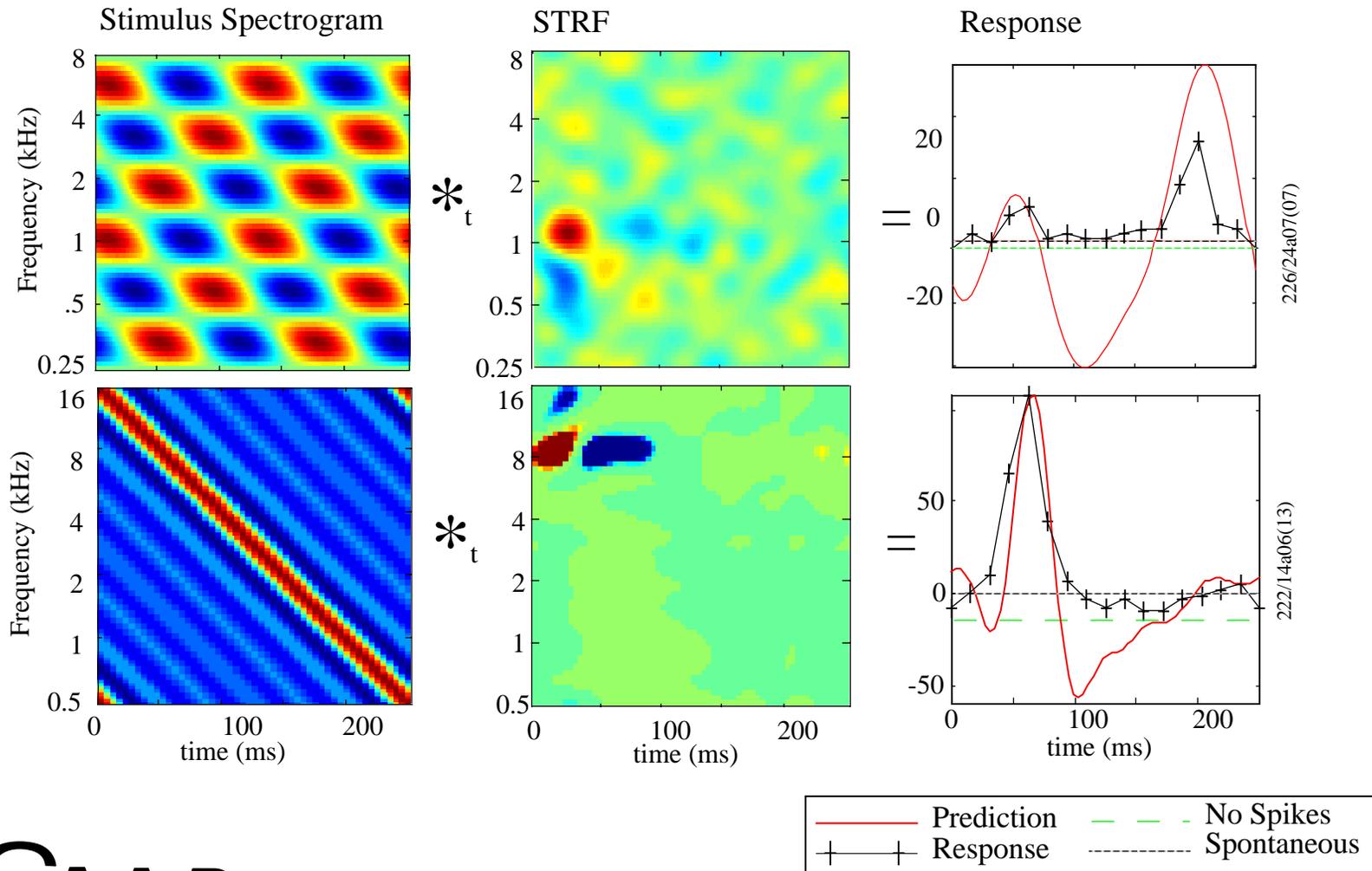
Linearity in Theory

Assuming linearity, the STRF predicts the response to any broadband dynamic stimulus, including single ripples moving in either direction (first two rows) and combinations of upward and downward moving ripples.



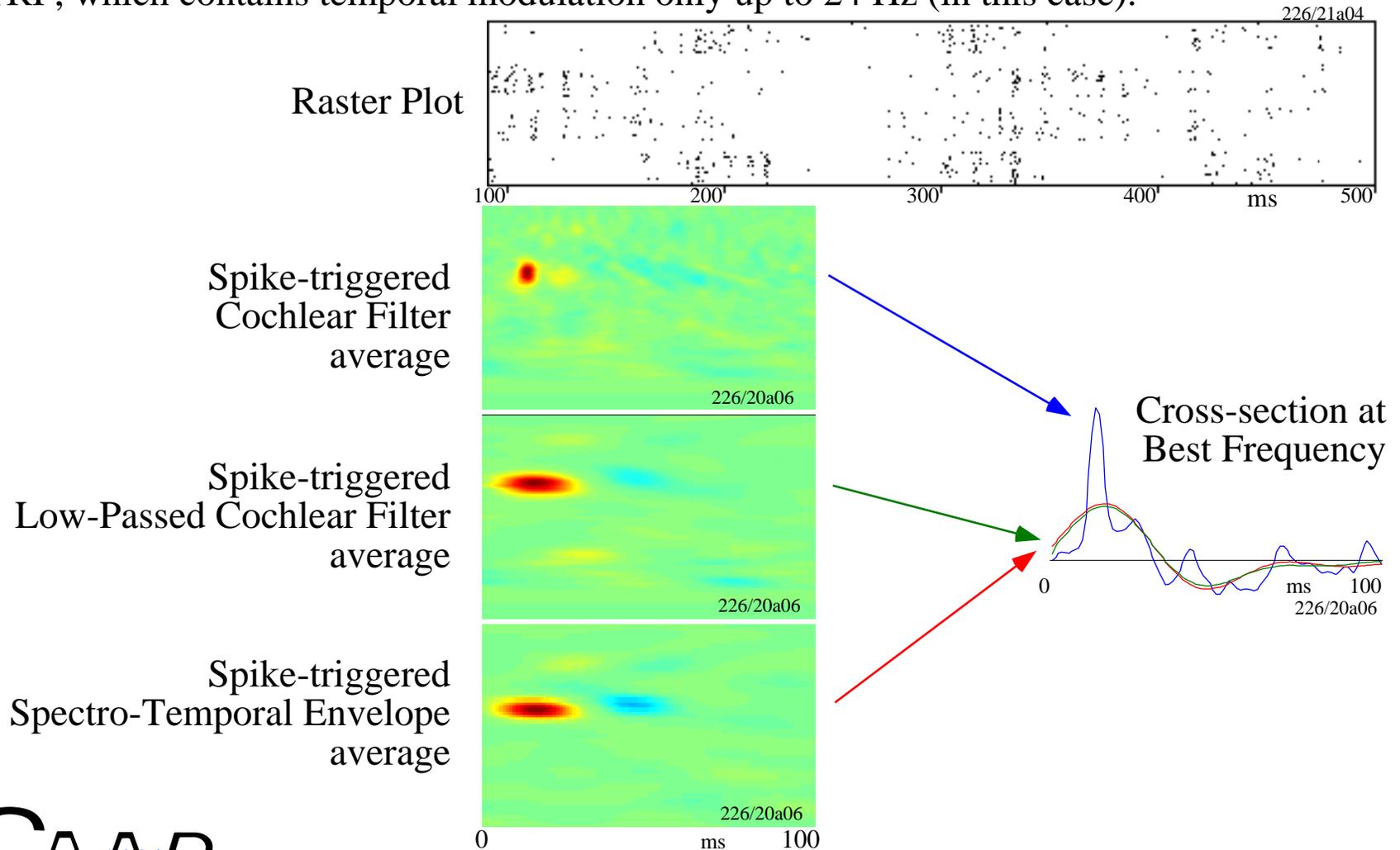
Linearity in Practice

The correlation between predicted and actual response is quite good for most cells. Since cells cannot fire at negative rates, any prediction should be half-wave rectified before comparing to the actual response.



Fast Responses

Some units respond well at time scales as fast as ~ 10 ms. This is seen both in the raster plot and in the STRF. When the output of the filter bank is low-passed at 25 Hz, the resulting STRF looks much more like the Spectro-Temporal Envelope generated STRF, which contains temporal modulation only up to 24 Hz (in this case).



Quadrant Separability

An STRF can fall into one of three categories:

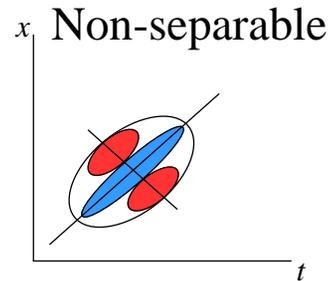
- **Non-separable:** The transfer function is an arbitrary (complex-conjugate symmetric) function of ripple frequency and ripple velocity.

- **Quadrant separable:** The transfer function *within each quadrant* is a product of a function of ripple frequency and a function of ripple velocity. The envelope of the STRF is a simple product of a function of spectrum and a function of time.

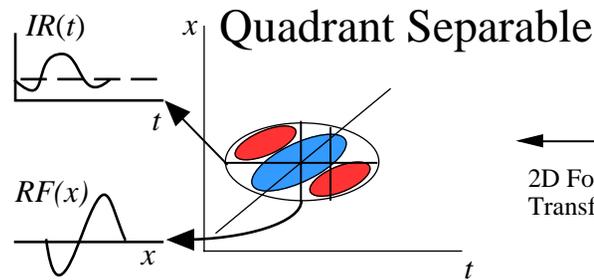
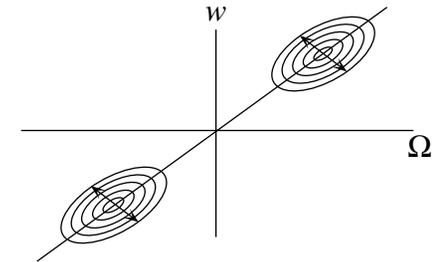
- **Fully separable:** The transfer function is the product of a function of ripple frequency and ripple velocity *everywhere*. The resulting STRF is a product of a function of spectrum and a function of time.

Spectro-Temporal Domain

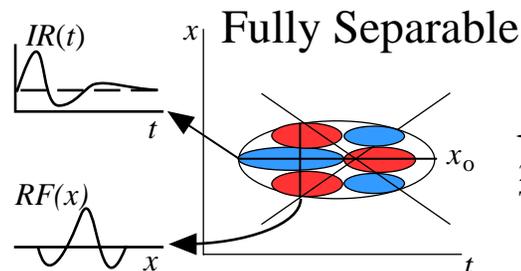
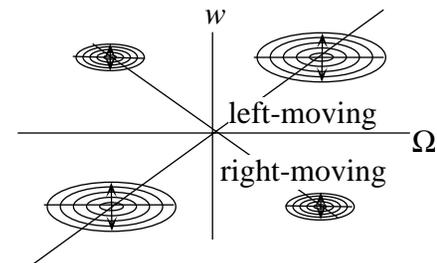
Fourier Domain



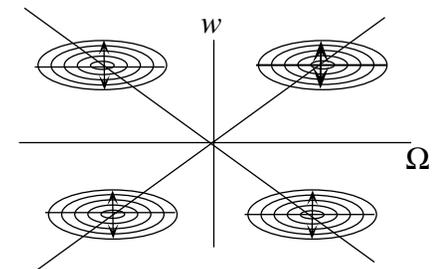
2D Fourier Transform



2D Fourier Transform

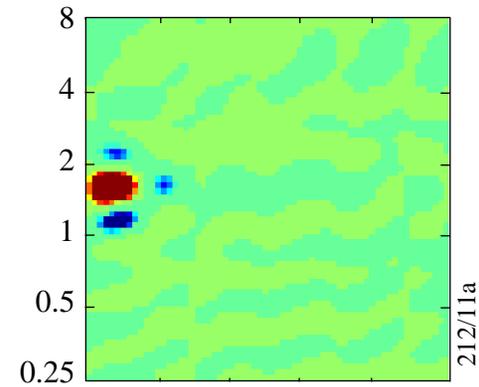
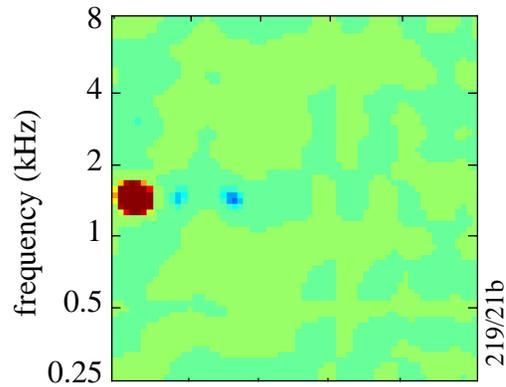


2D Fourier Transform

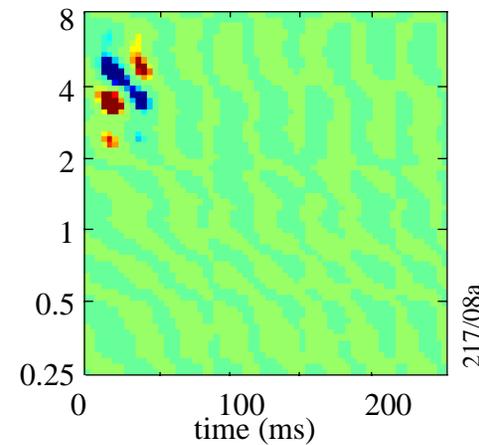
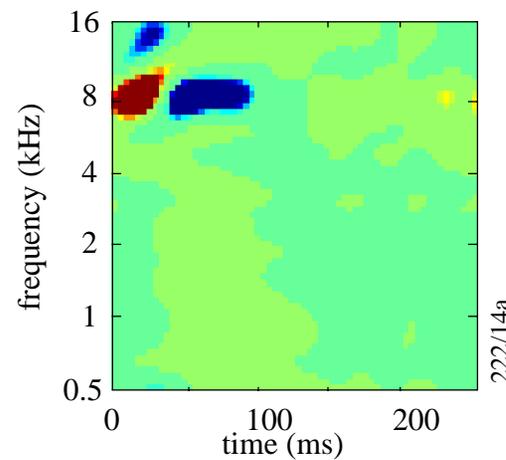


Separability in STRFs

Examples of Experimentally obtained STRFs



Separable



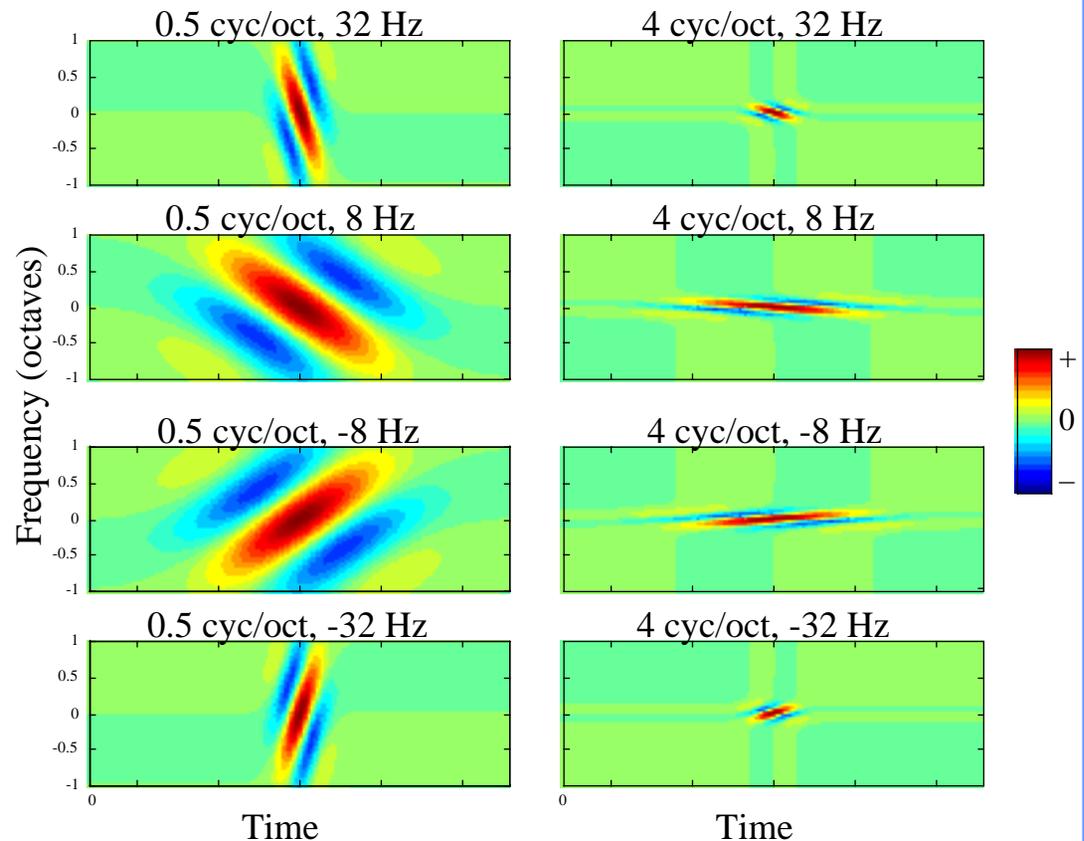
Non-separable

Note the variety of spectral and temporal behaviors

Cortical Filter Model

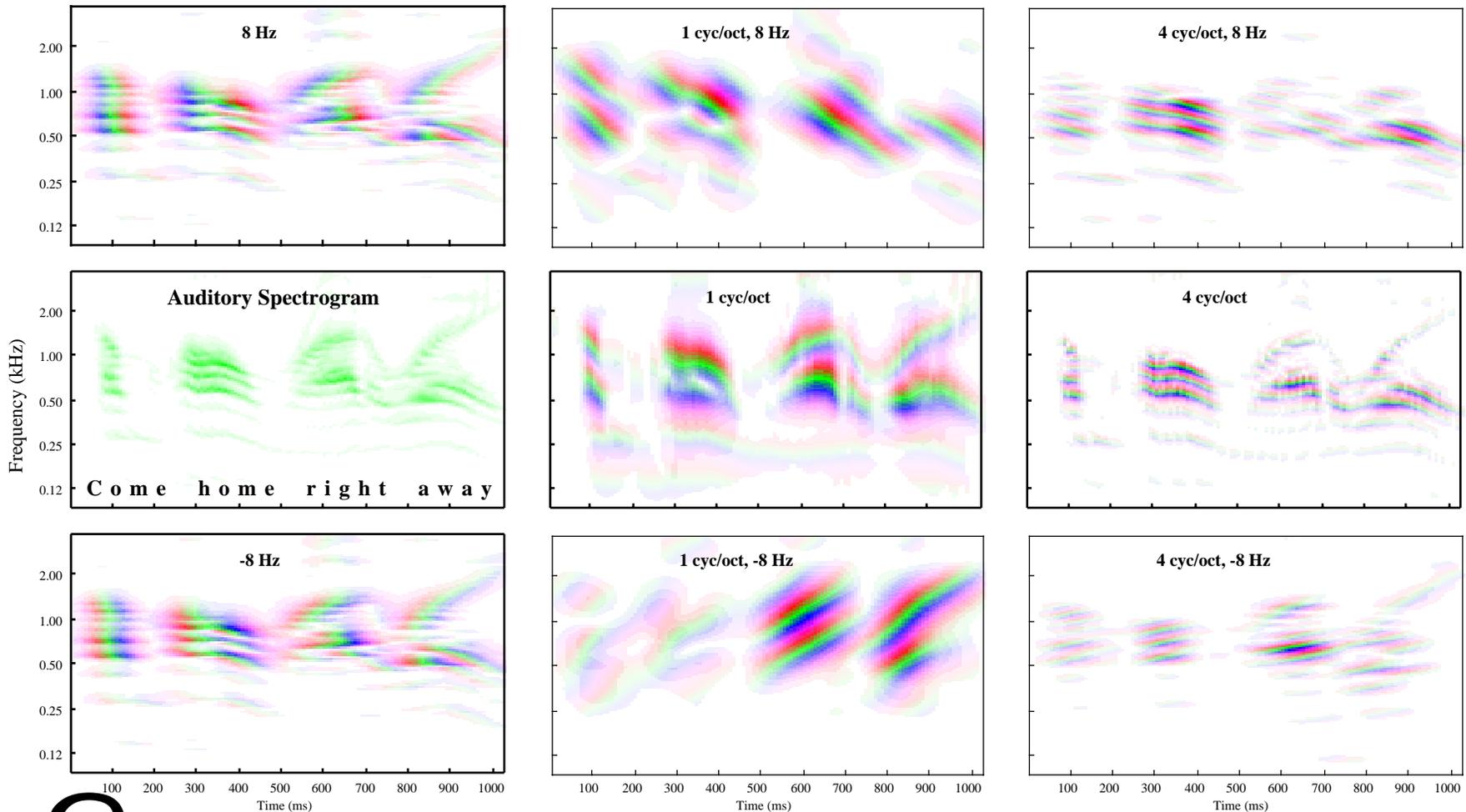
- Response fields in AI have characteristic shapes both spectrally and temporally.
- AI cells respond well only to a small set of moving ripples around a particular spectral peak spacing and velocity.
- We find cortical cells with all center frequencies, spectral symmetries, bandwidths, latencies and temporal impulse response symmetries.
- Therefore AI decomposes the input spectrum into different spectrally and temporally tuned channels.
- Equivalently, a population of cells, tuned around different moving ripple parameters, can effectively represent the input spectrum at multiple scales.

Theoretical ripple filters used to generate a 'cortical representation'



The Cortical Representation

Spectrally narrow cells pick out the fine features of the spectral profile, whereas broadly tuned cells pick out the coarse outlines of the spectrum. Similarly, dynamically sluggish cells will respond to the slow changes in the spectrum, whereas fast cells respond to rapid onsets and transitions. In this manner, AI is able to encode multiple views of the same dynamic spectrum.



Selected References

Spectro-Temporal Averaging Methods

- ❑ Calhoun BM, Miller RL, Wong JC, and Young ED, 11th International Symposium on Hearing (1997).
- ❑ Eggermont JJ, Hearing Research 66 (1993) 177-201.

Dynamical Transfer Function papers

- ❑ Kowalski NA, Depireux DA and Shamma SA, J.Neurophys. 76 (5) (1996) 3503-3523, and 3524-3534.
- ❑ Depireux DA, Simon JZ and Shamma SA, Comments in Theoretical Biology (1997).

Stationary Transfer Function papers

- ❑ Shamma SA, Versnel H and Kowalski NA, J. Auditory Neuroscience (1) (1995) 233-254, and 255-270, and 271-285.
- ❑ Schreiner CE and Calhoun BM. Auditory Neurosci., 1 (1994) 39-61.

Related analysis techniques and models

- ❑ Wang K and Shamma SA, IEEE Trans. on Speech and Audio 2(3) (1994) 421-435, and 3(2) (1995) 382-395.
- ❑ Shamma SA, Fleshman JW, Wiser PR and Versnel H, J. Neurophys 69(2) (1993) 367-383.