Representation of Complex Spectra in Auditory Cortex

Shihab Shamma

Didier Depireux

Jonathan Simon

Institute for Systems Research
University of Maryland

Work supported by a MURI grant from the Office of Naval Research, a training grant from the National Institute on Deafness and Other Communication Disorders, and the National Science Foundation

Abstract

Natural sounds are broadband and dynamic. To understand their encoding in primary auditory cortex (AI), we have characterized the responses of units in AI with elementary versions of such spectra—moving ripples. Ripples are broadband sounds with a sinusoidal envelope along the log frequency axis, that move up or down with a constant velocity. Speech spectra can be decomposed into a superposition of ripples with different densities and velocities.

If AI units are linear, then it is possible to predict how a unit responds to any broadband dynamic stimulus by first measuring its responses to all elementary ripples (i.e., measure the ripple transfer function), and then superposing the responses to these ripples, each according to its weight in the input. We have successfully demonstrated the linearity of AI units in the past using ripples either stationary or moving only downward in frequency. The data described in this poster will show that transfer functions are also separable for up-moving ripples, but that the two transfer functions may well be different. Hence AI units are not always fully separable, but only separable by quadrant. We shall discuss the implications of these results and show examples of predicted and measured responses to speech.

Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Summary

Natural sound
- loudness (≈ intensity)
- pitch (≈ tonal height)
- timbre (the rest, e.g. spectral envelope)

Question: How is timbre encoded in primary auditory cortex?

Important Concepts:
- *Response Field* (RF): range of frequencies that influence a neuron. RF is a function of time.
- *Ripples*: broadband sounds with sinusoidally modulated spectral envelope.
- Data analysis based on *linear systems*; by varying ripple frequency and velocity, we measure the transfer function. The inverse Fourier transform gives the *spectro-temporal* RF (STRF).

We show predictions of single-unit responses to complex spectra, including:
- ***Linearity*** of responses to dynamic ripples: responses to upward and downward moving ripples can be superimposed to predict responses to arbitrary combinations.
- ***Separability*** of spectral and temporal measurements of the responses: spectral properties can be measured independently of temporal properties in some cases.

We find:
- Cells can be characterized by an STRF, separable or non-separable.
- Cells behave like a linear system: when presented with a sum of several profiles, the response is the sum of the responses to the individual profiles.

We conclude that the combined spectro-temporal decomposition in AI is an affine wavelet transformation of the input, in concert with a similar temporal decomposition. The auditory profile is the result of a multistage process which occurs early in the pathway. This pattern is projected centrally where a multiscale representation is generated in AI by STRFs with a range of widths, asymmetries, BFs, time lags and directional sensitivities.

Center for Auditory and Acoustic Research

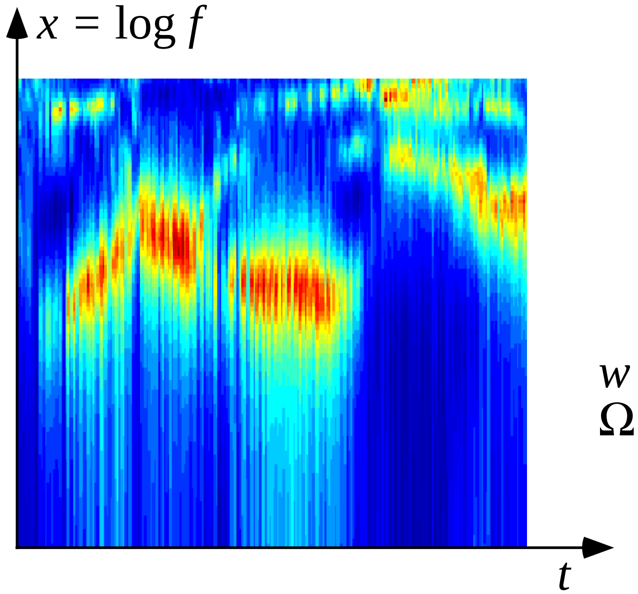
Institute for Systems Research
University of Maryland

Spectro-Temporal Transform

- Frequencies are mapped along the cochlea on a log (frequency) axis.
- Since natural sounds are dynamic, we need a time axis.
- Therefore we use two-dimensional functions of log(frequency) and time.

- Consider the (Fourier) space dual to the two-dimensional spectro-temporal space.
- For linear systems, the spectro-temporal domain and its Fourier domain are equivalent. Analysis is often conceptually simpler in the Fourier domain.

Spectrogram envelope of a speech fragment '*water all year*'



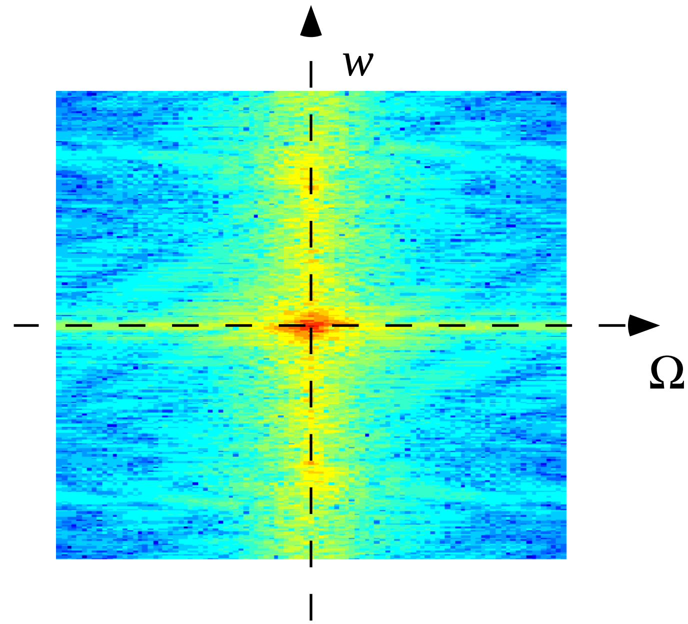
Fourier Transform

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

Inverse Transform

w = "ripple velocity"
Ω = "ripple frequency"

Fourier transform of the envelope of the spectrogram



Real functions in the spectro-temporal domain give rise to complex conjugate symmetric functions in the Fourier domain.

Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Spectro-Temporal Response

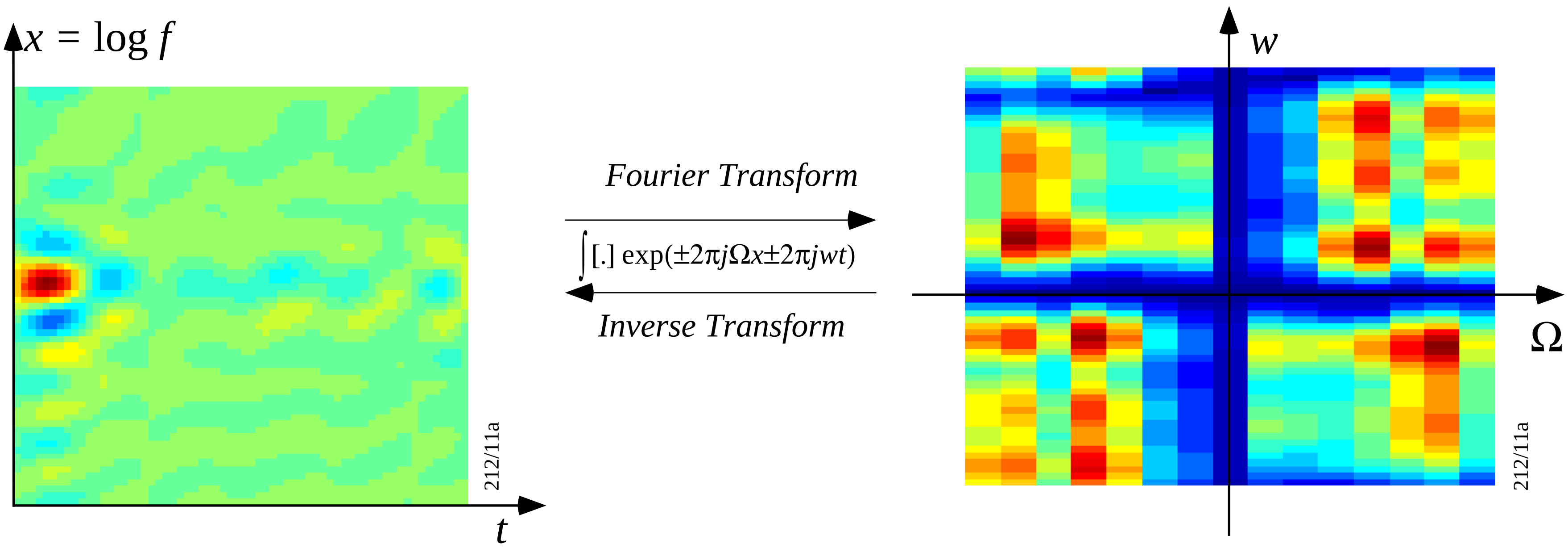
- The spectro-temporal response field of a neuron is the usual response field made time-dependent. Equivalently, it is the temporal impulse response for each frequency.

- 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

2 Dimensional Transfer Function of the same neuron



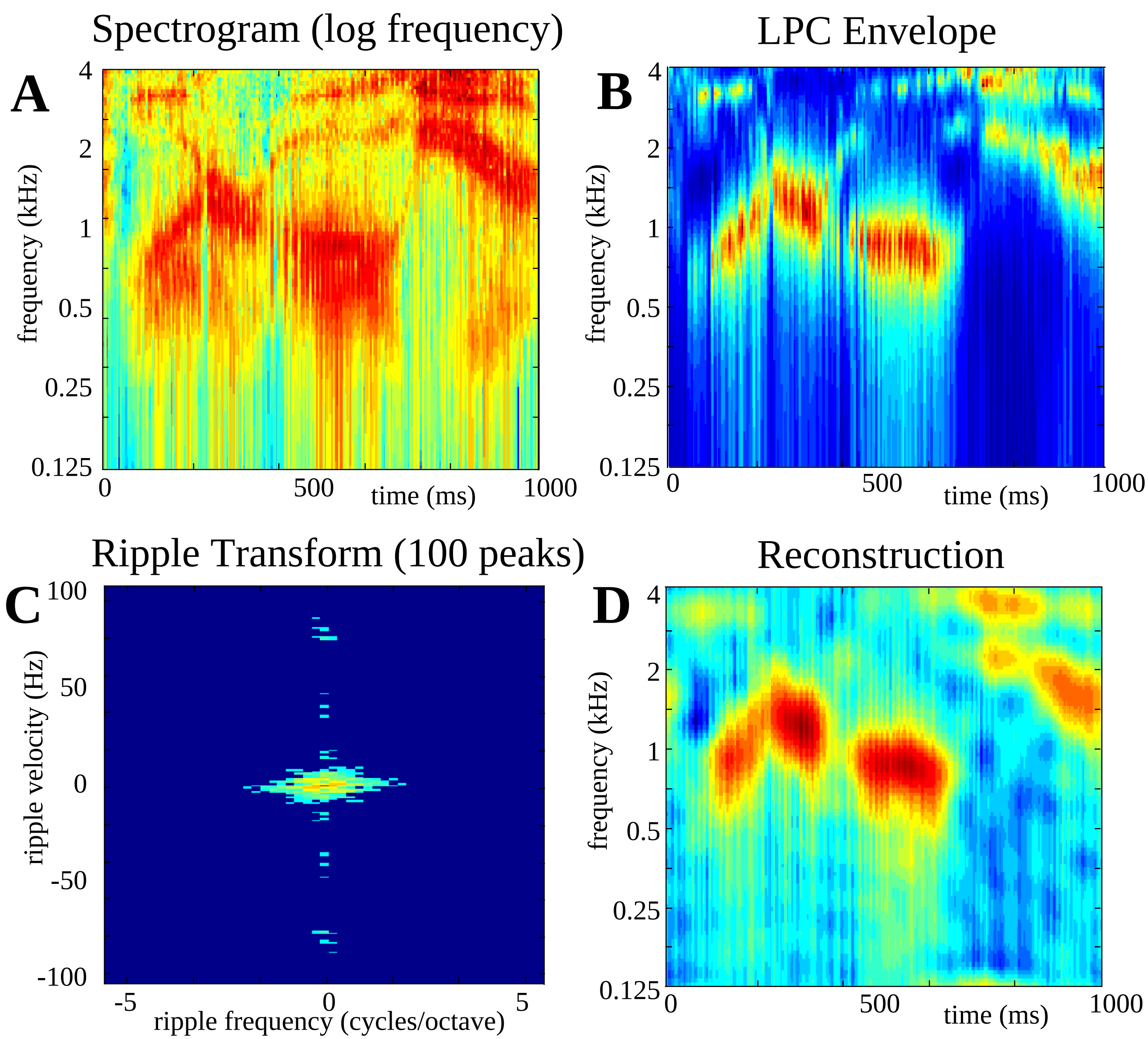
Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Ripple Decomposition

Ripple decomposition of a broadband dynamic sound

(A) The envelope of a speech fragment is Fourier transformed in (B). 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.



"*water all year*"

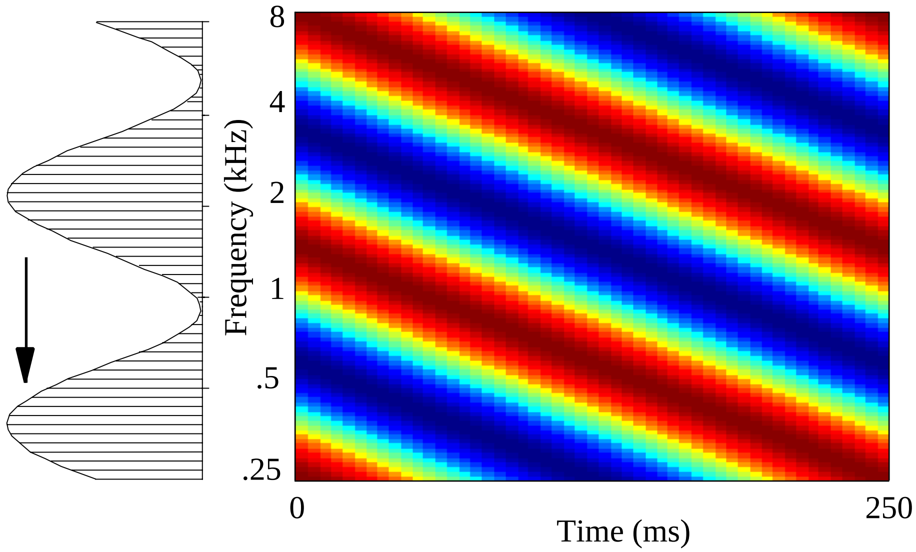
Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

The Ripple Stimulus

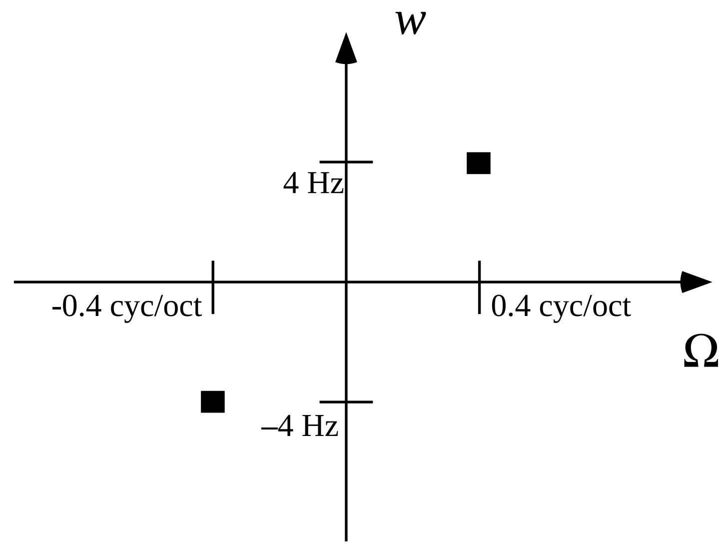
Ripples are broadband sounds with sinusoidally modulated spectral envelope along the log (frequency) axis, analogous to visual gratings.

Ripple in Spectro-Temporal Space (Spectrogram)

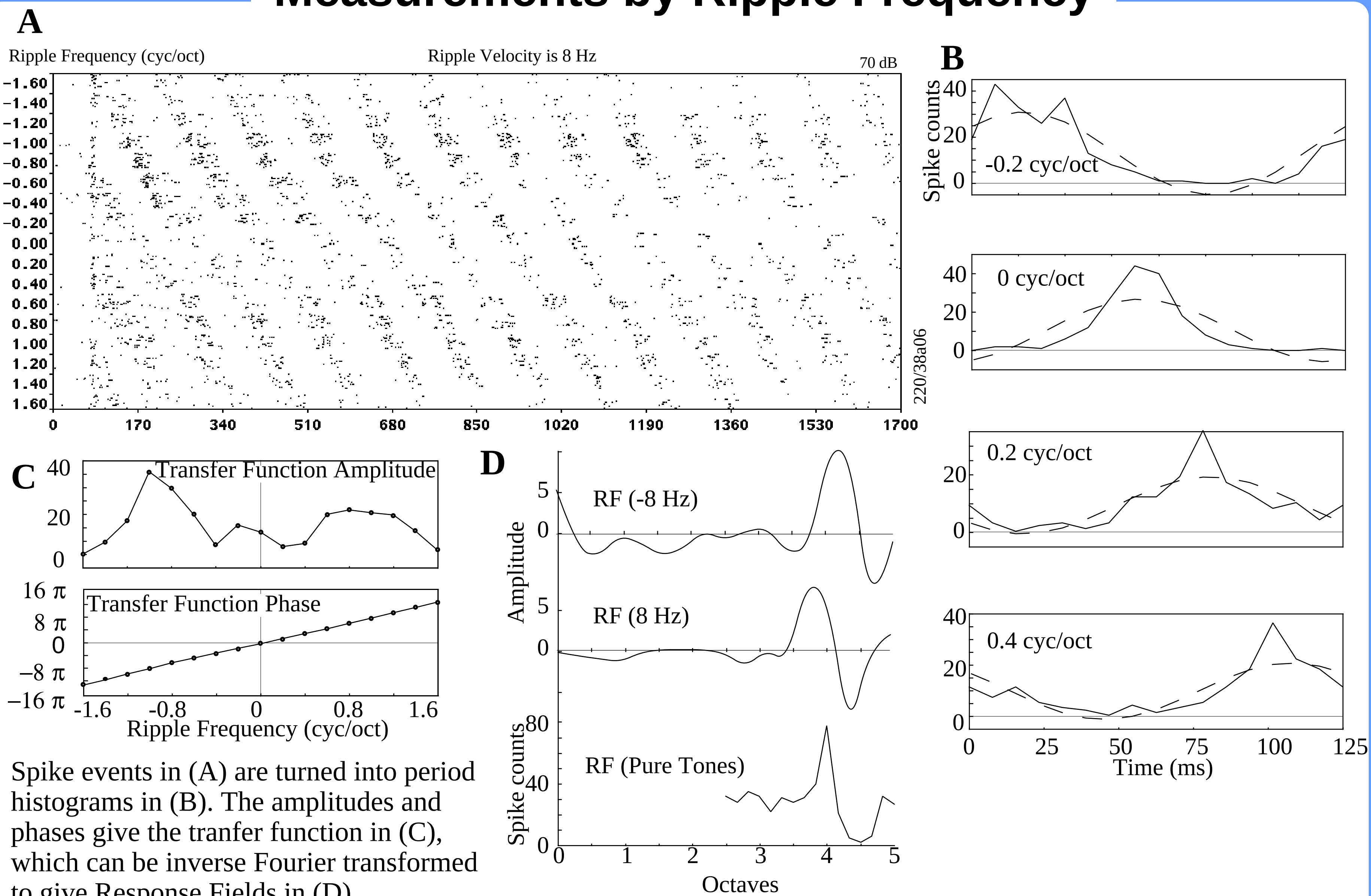


The Fourier transform of a “ripple” has support only on a single point (and its complex conjugate).

Ripple in Fourier Space



Measurements by Ripple Frequency

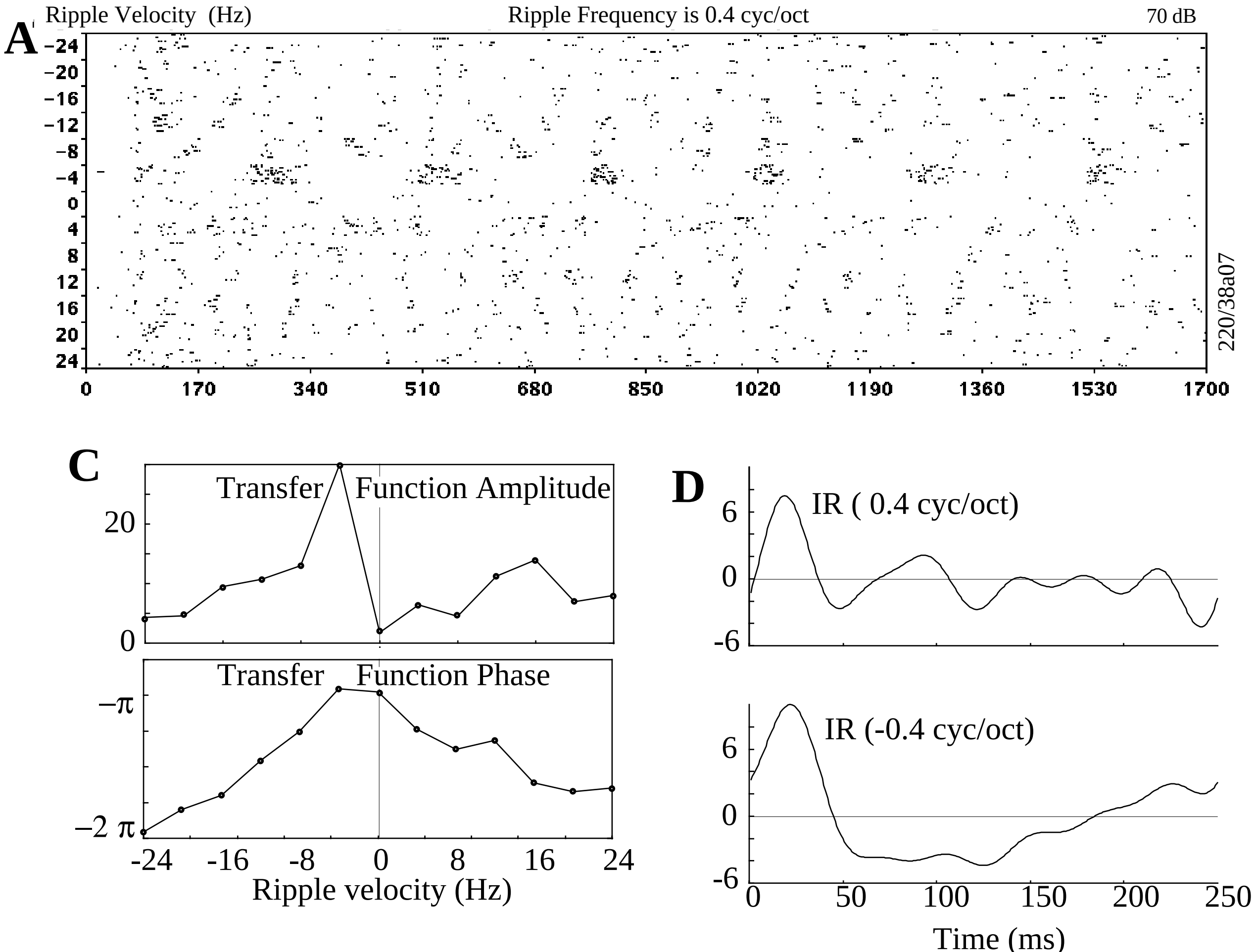


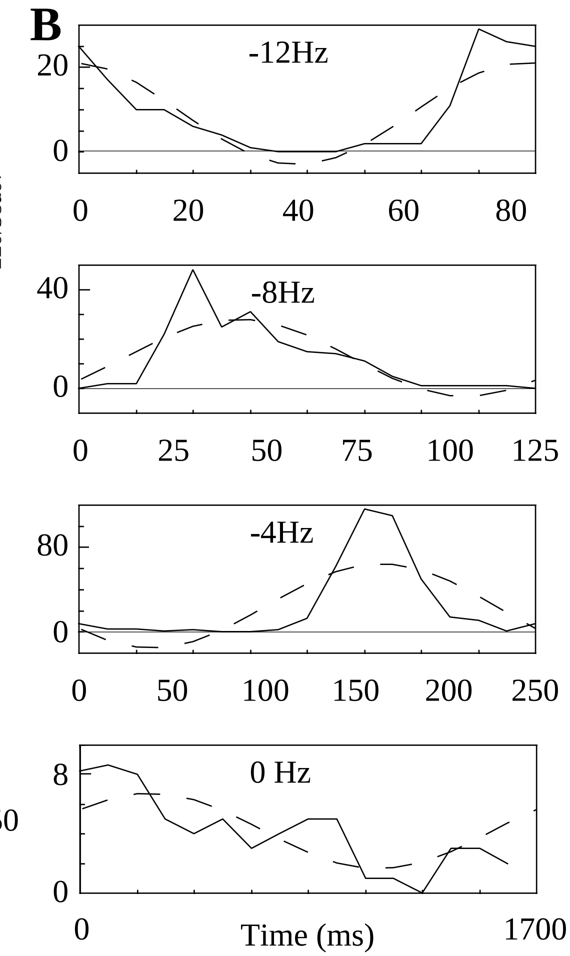
Spike events in (A) are turned into period histograms in (B). The amplitudes and phases give the tranfer function in (C), which can be inverse Fourier transformed to give Response Fields in (D).

Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Measurements by Ripple Velocity





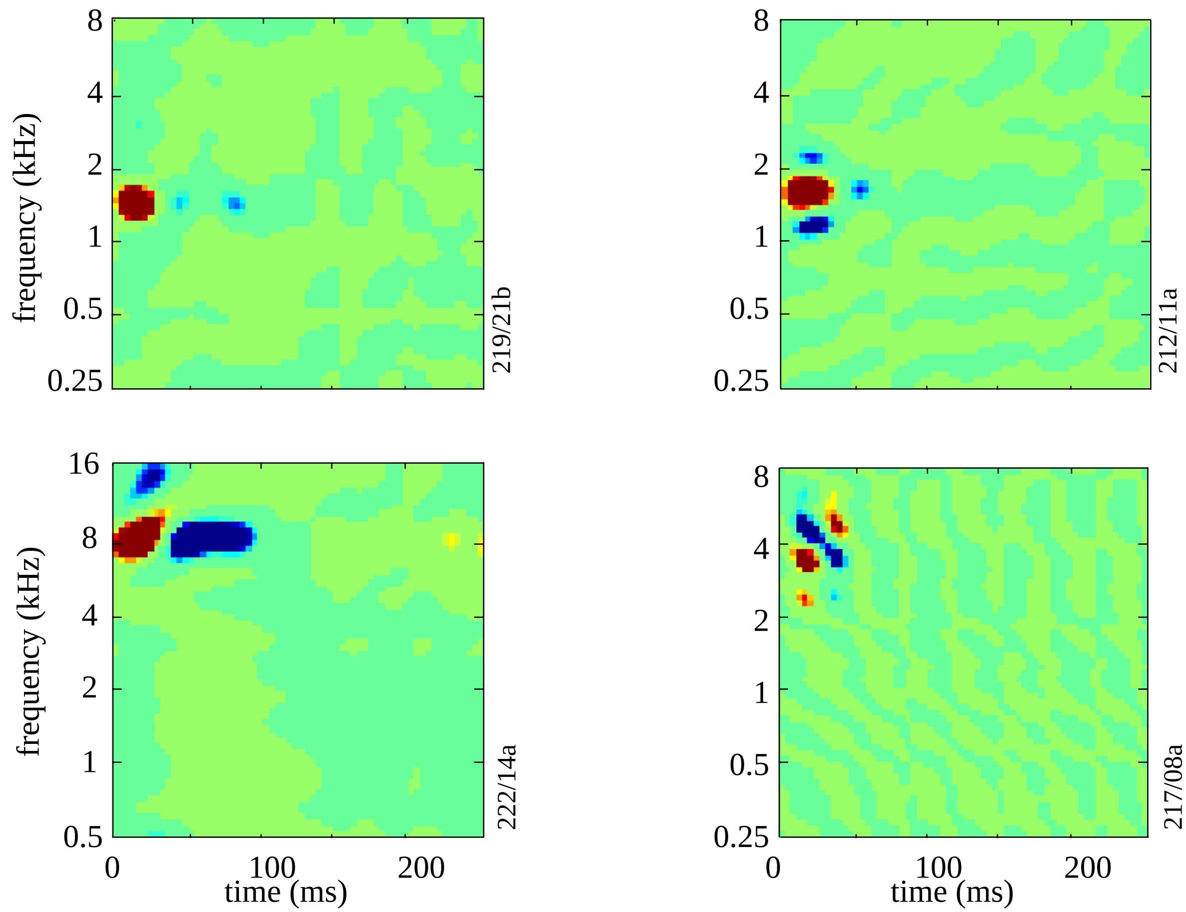
Spike events in (A) are turned into period histograms in (B). The amplitudes and phases give the tranfer function in (C), which can be inverse Fourier transformed to give Impulse Responses in (D).

Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Spectro-Temporal Response Fields

Examples of Experimentally obtained STRFs



Separable

Non-separable

Note the variety of spectral and temporal behaviors

Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

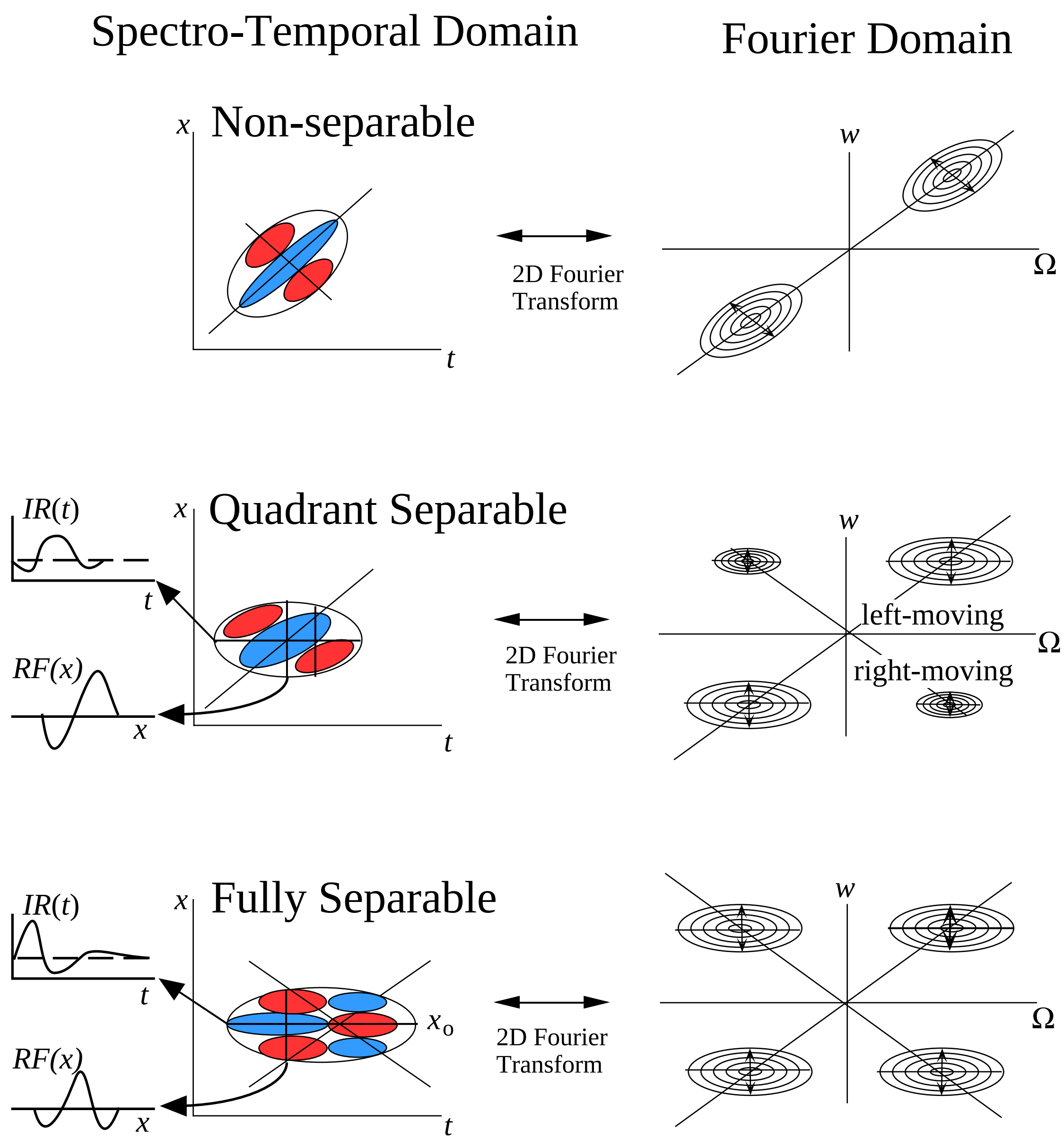
Quadrant Separability

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 freqency and ripple velocity *everywhere*. The resulting STRF is a product of a function of spectrum and a function of time.

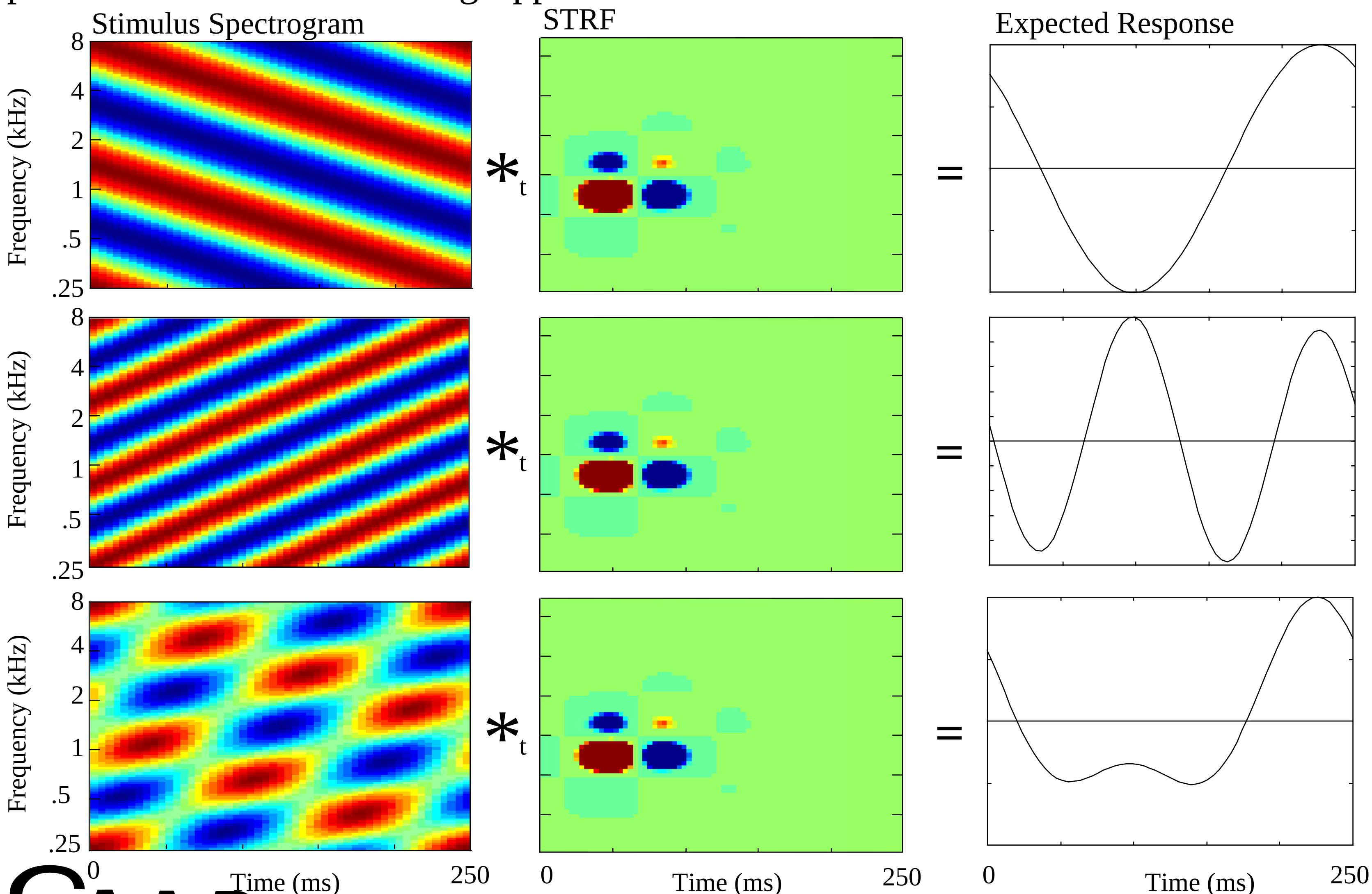


Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

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.

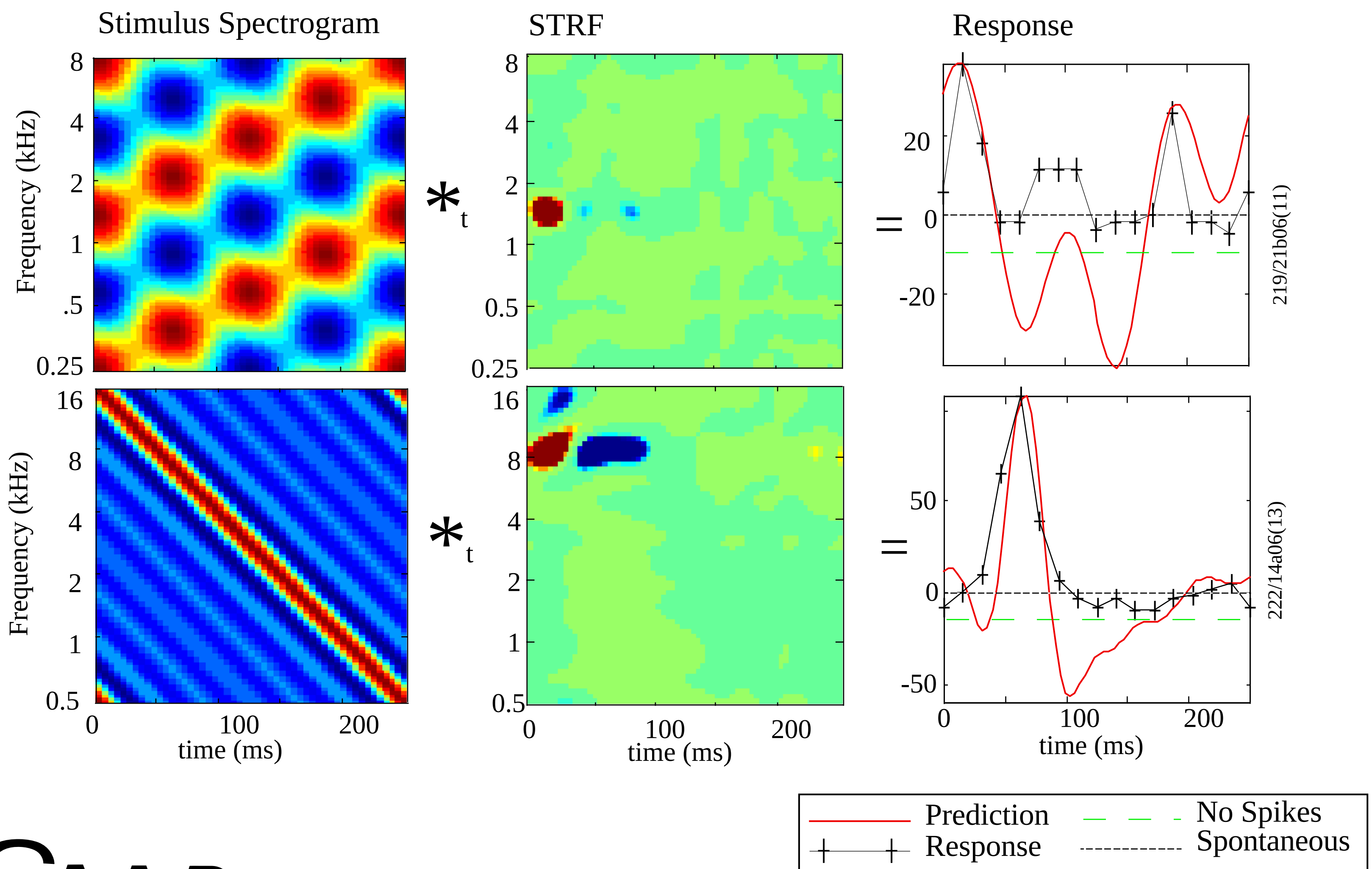


Center for Auditory
and Acoustic Research

Institute for Systems Research
University of Maryland

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.



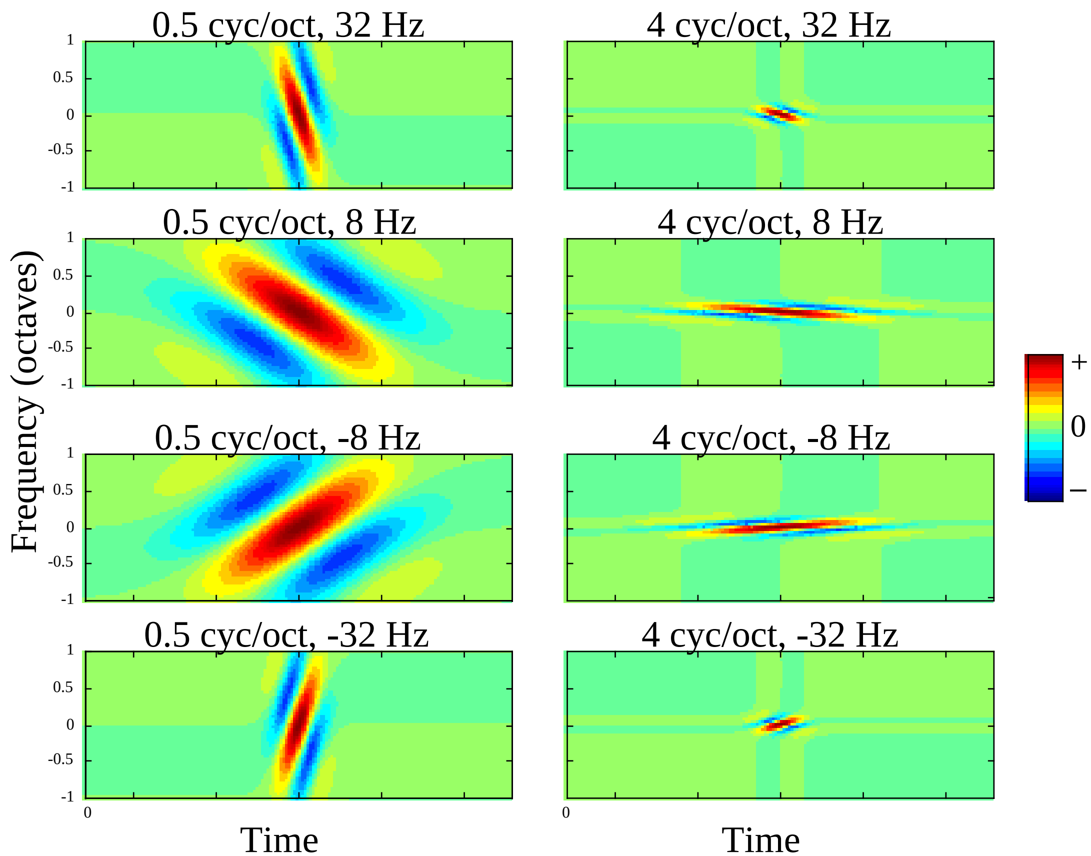
Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

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'

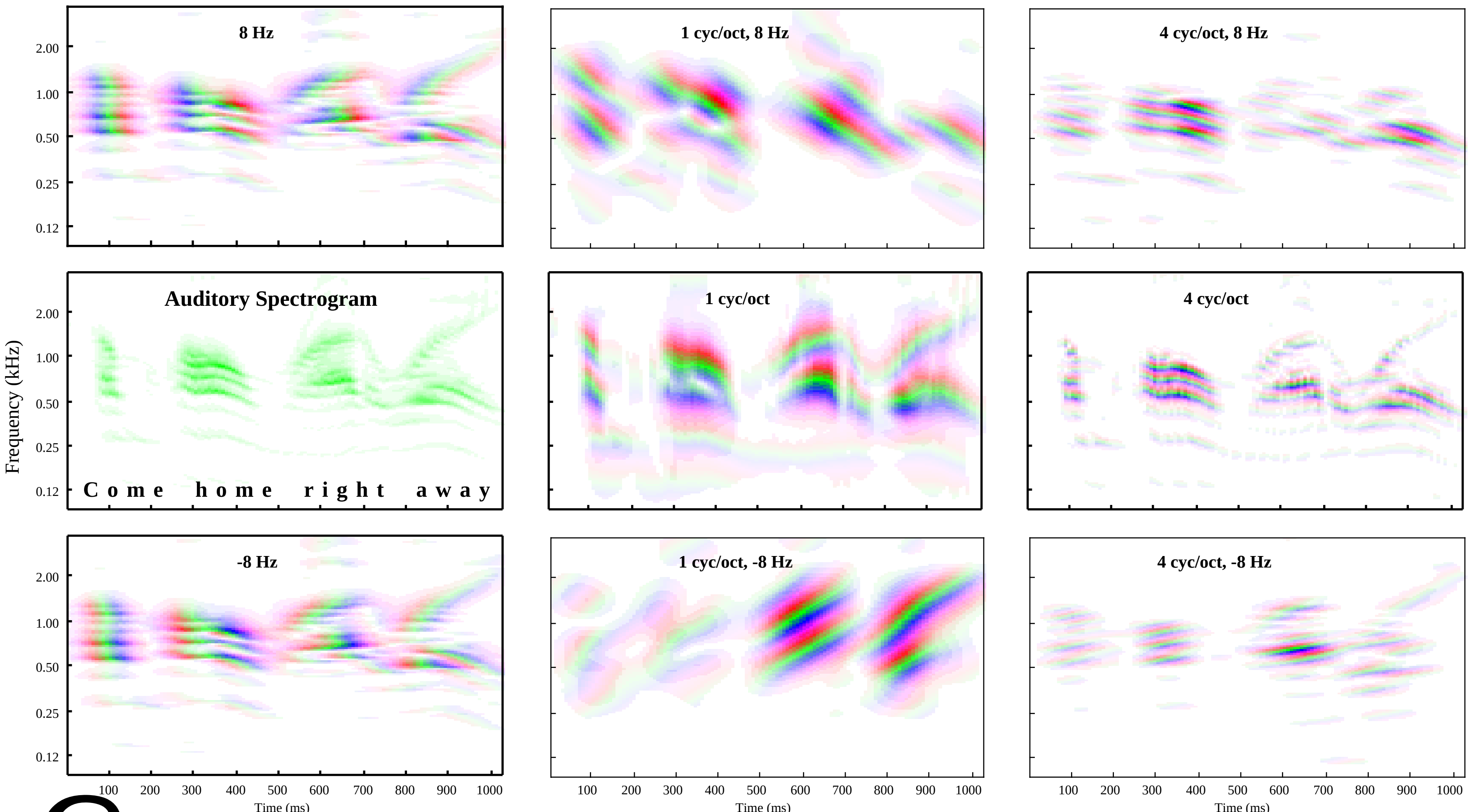


Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

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.



Center for Auditory and Acoustic Research

Institute for Systems Research
University of Maryland

Selected References

Dynamical 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 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.