COMP 546

Lecture 22

Spectrograms (revisited), Auditory filters

Thurs. April 5, 2018

Spectrogram

Partition a sound signal into B blocks of T samples each (i.e. the sound has BT samples in total).

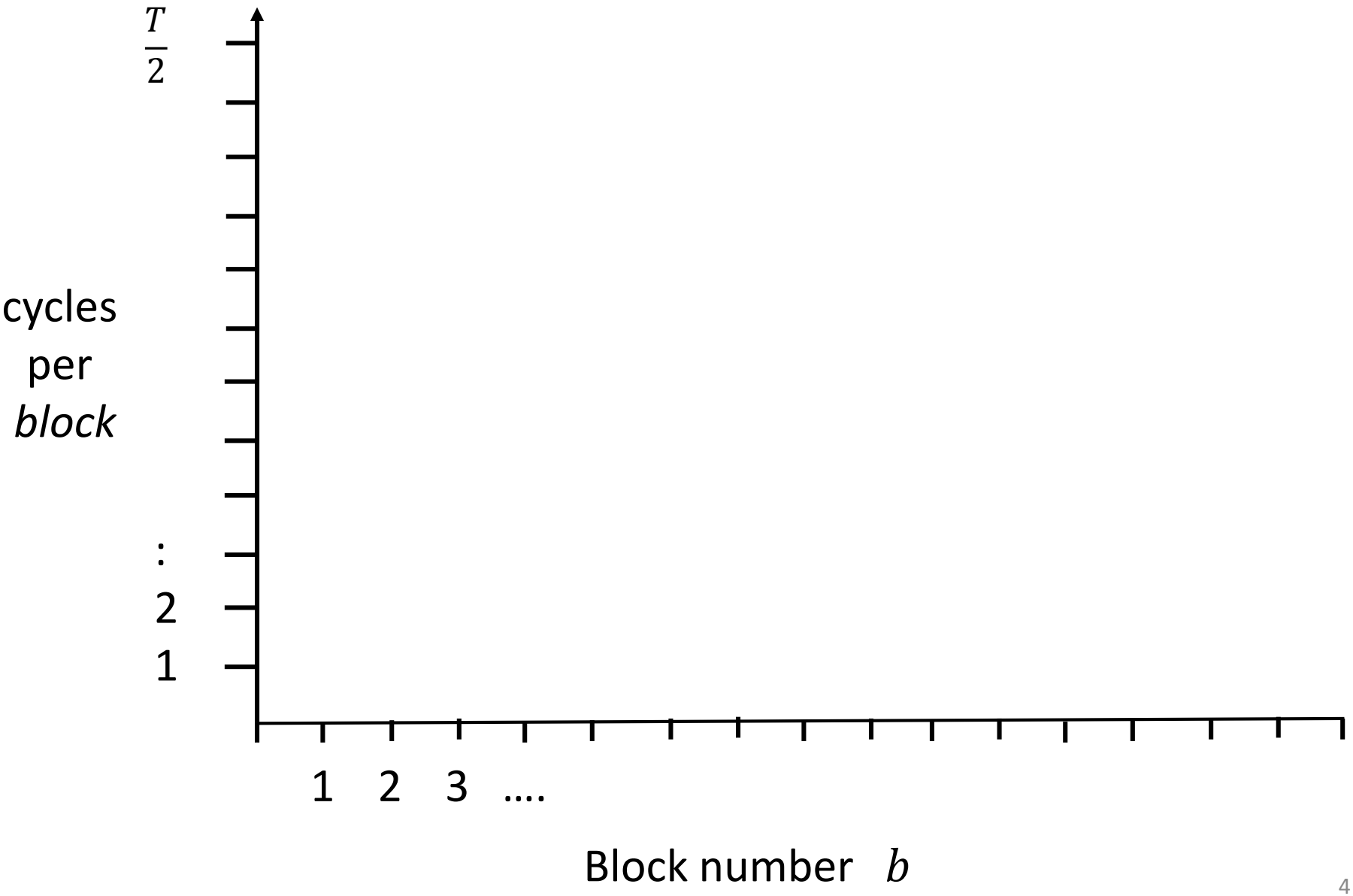
Spectrogram

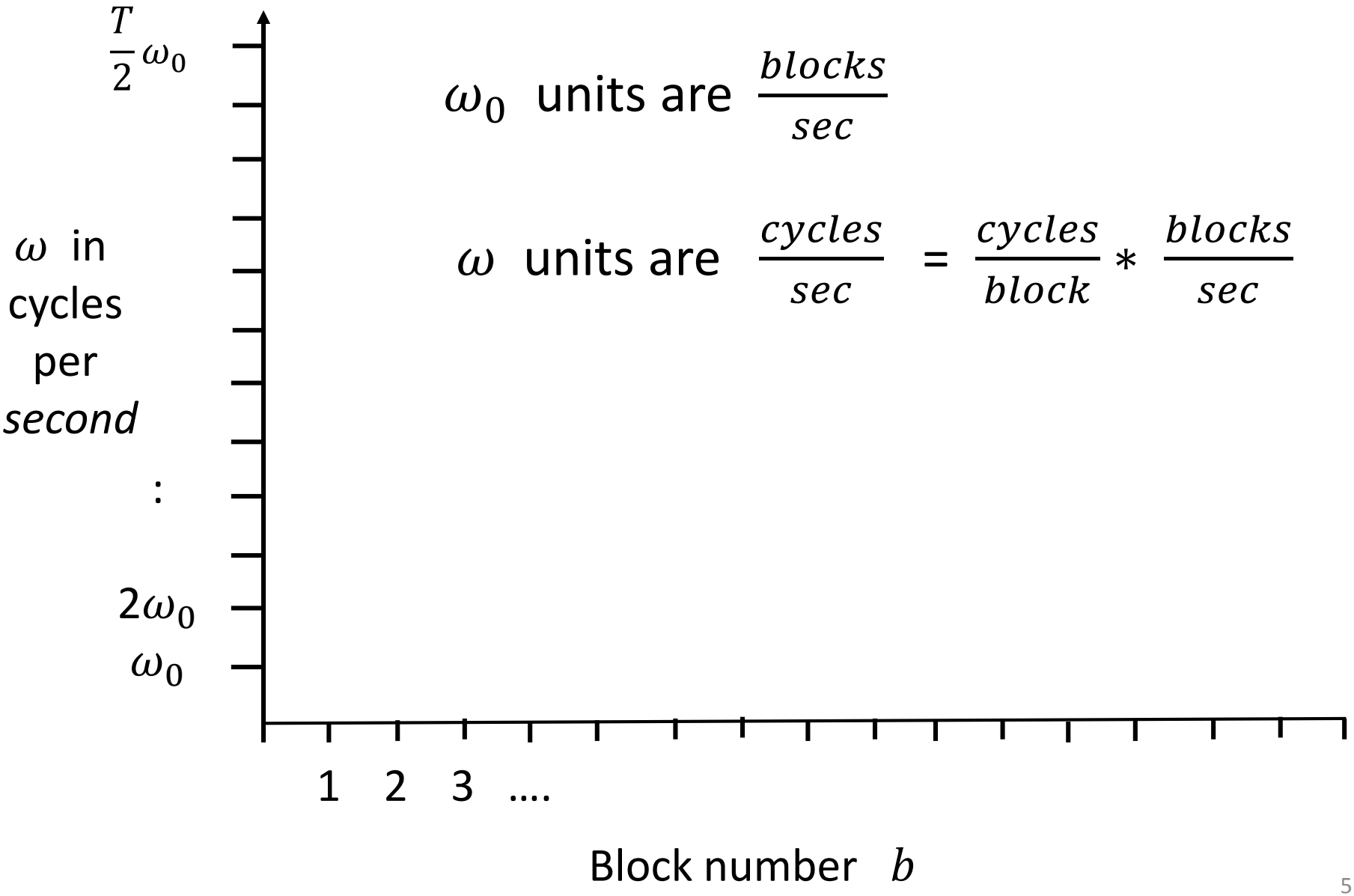
Partition a sound signal into B blocks of T samples each (i.e. the sound has BT samples in total).

Take the Fourier transform of each block.

Let b be the block number, and ω units be cycles per block.
[I will convert ω to cycles per second a few slides from now.]

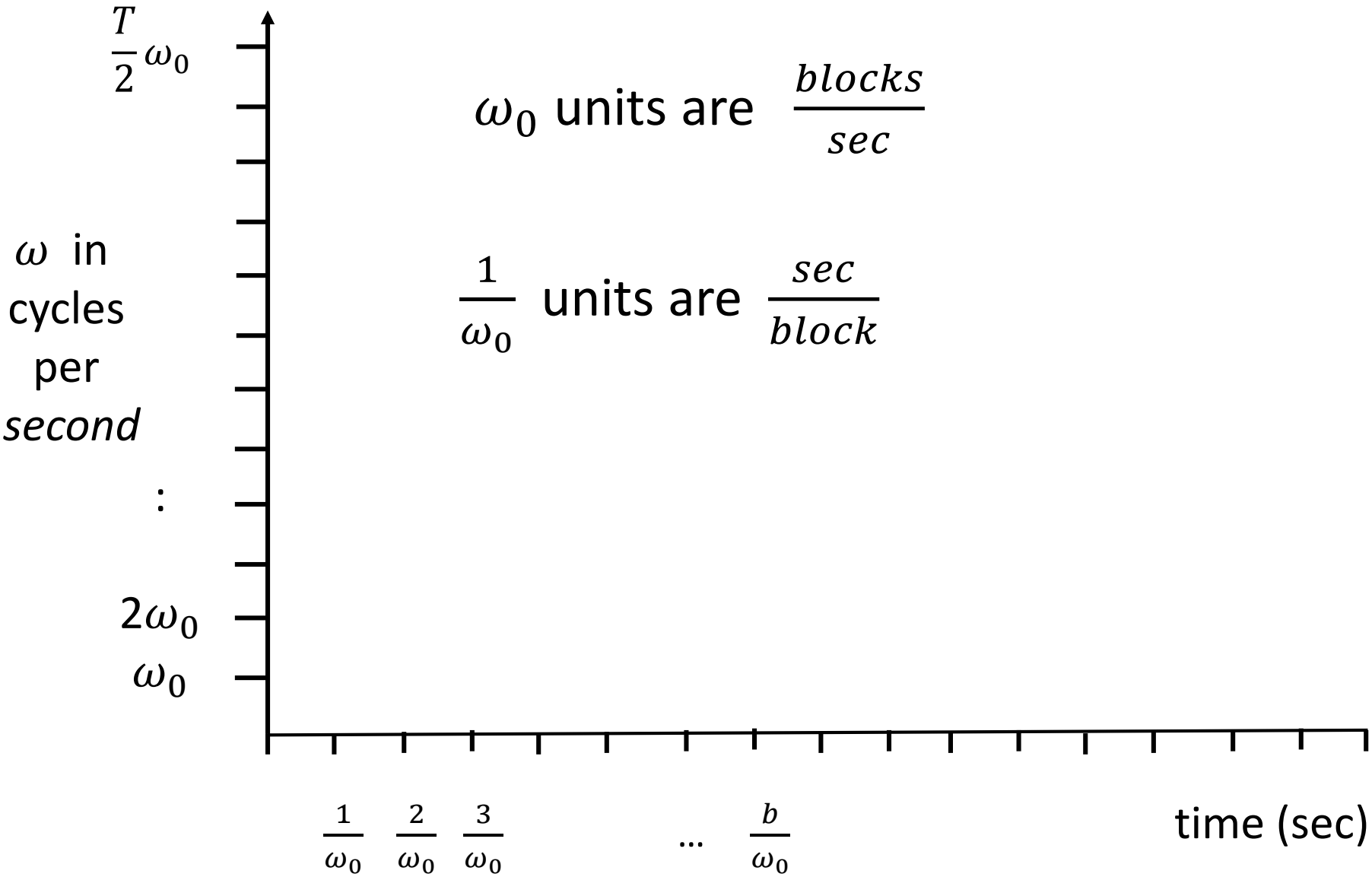
$$\hat{I}(b,\omega) \;=\; \sum_{t=0}^{T-1} I(\, b\,T + t)\; e^{-i\frac{2\pi}{T}\omega t}$$

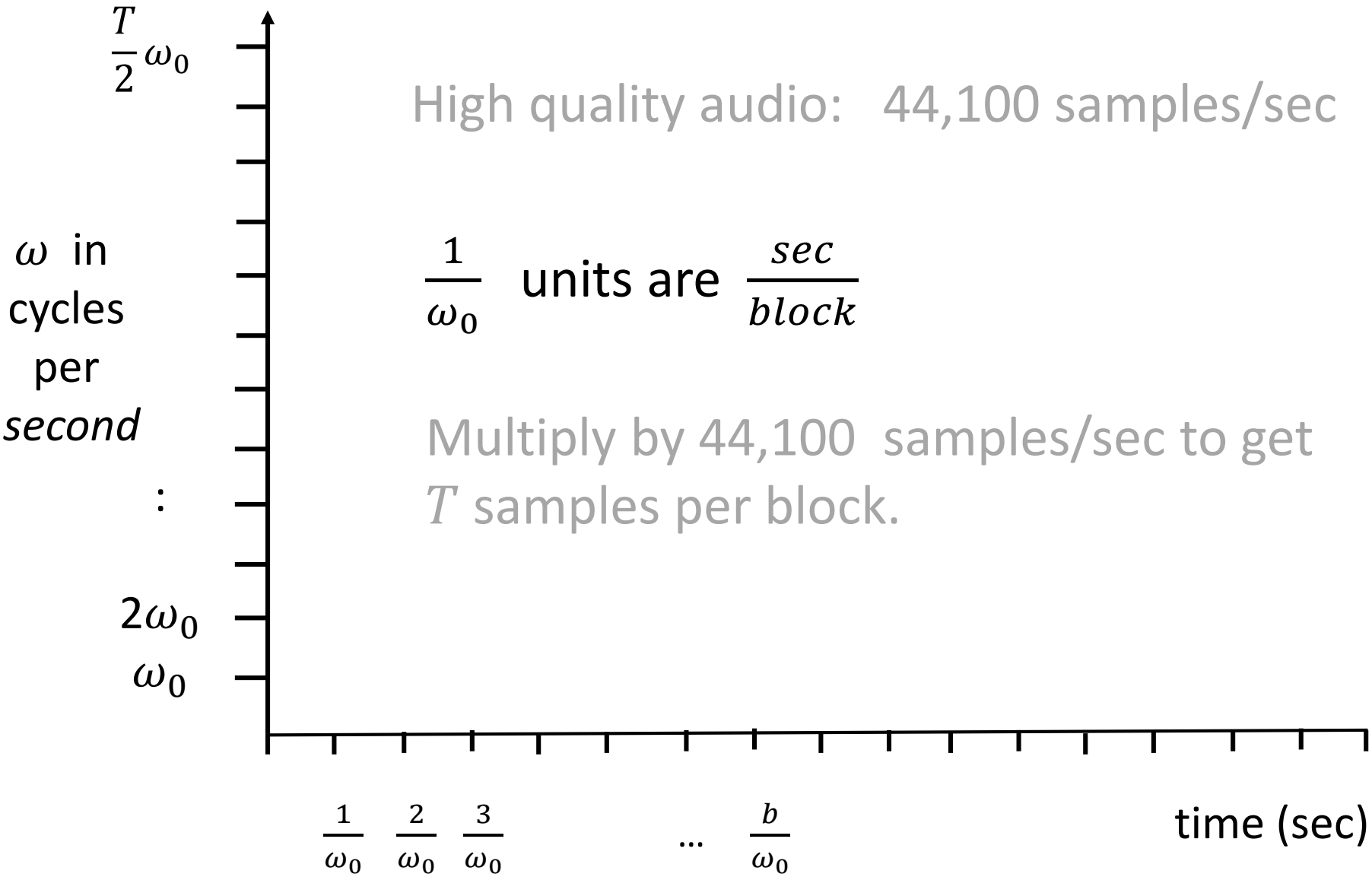

$\frac{T}{2}$
cycles
per
block
:
2
1
1 2 3 ….
Block number b



ω_0 units are $\frac{blocks}{sec}$

ω units are $\frac{cycles}{sec} = \frac{cycles}{block} * \frac{blocks}{sec}$

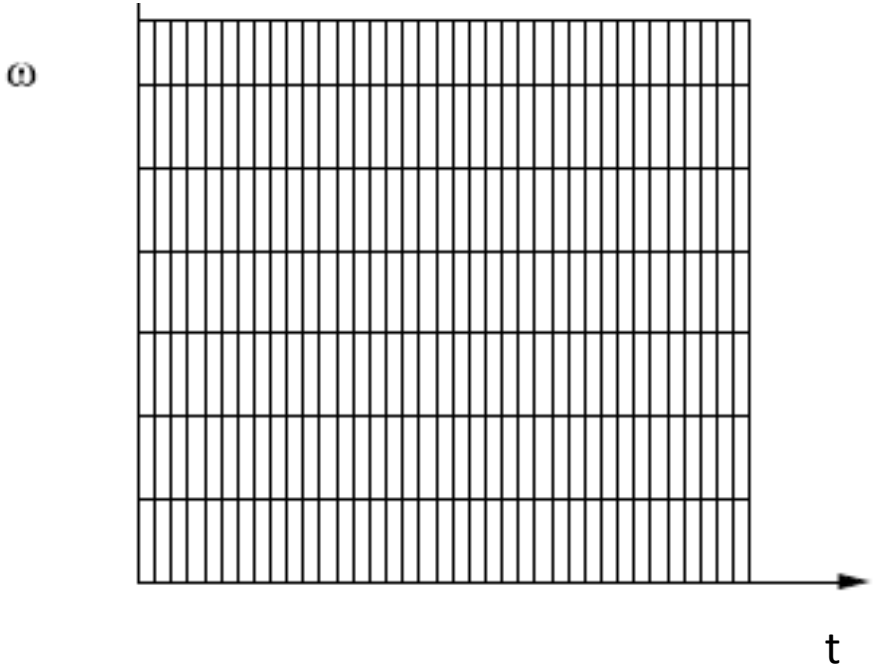

$\frac{T}{2}\omega_0$
ω in cycles per *second*
:
$2\omega_0$
ω_0
ω_0 units are $\frac{blocks}{sec}$
$\frac{1}{\omega_0}$ units are $\frac{sec}{block}$
$\frac{1}{\omega_0}$
$\frac{2}{\omega_0}$
$\frac{3}{\omega_0}$
…
$\frac{b}{\omega_0}$
time (sec)



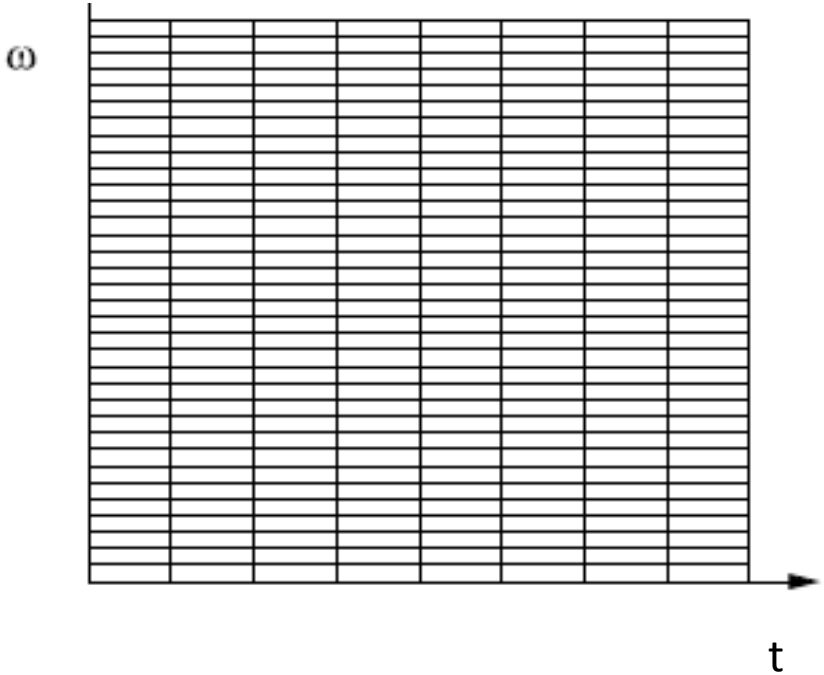
High quality audio: 44,100 samples/sec

$\frac{1}{\omega_0}$ units are $\frac{sec}{block}$

Multiply by 44,100 samples/sec to get T samples per block.



T smaller

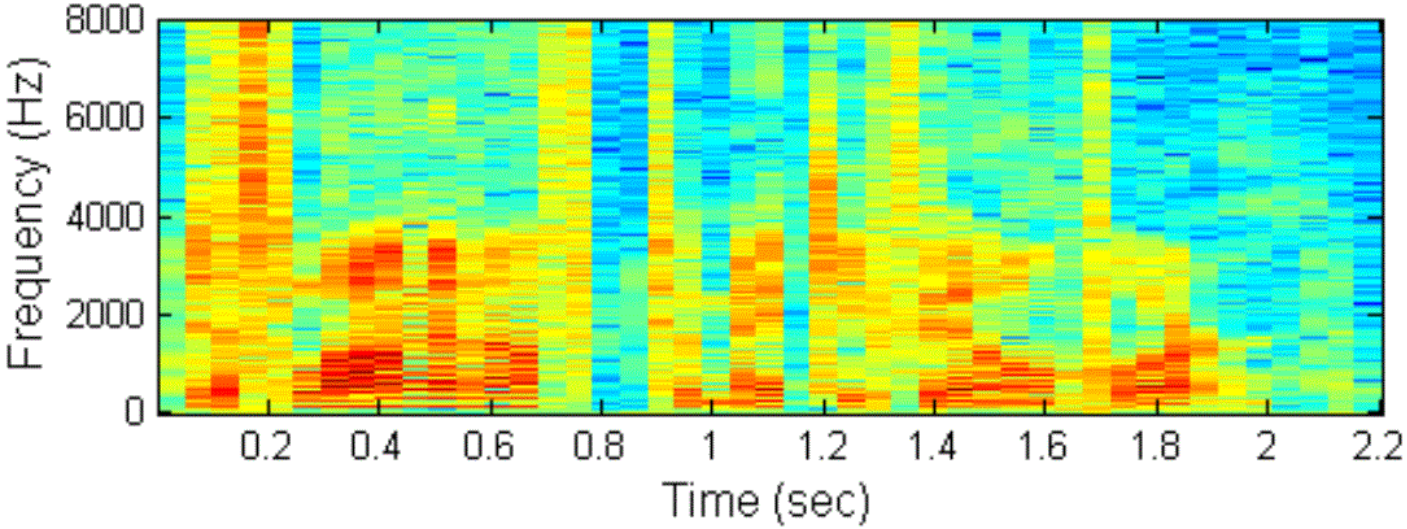


T larger

e.g. T = 512 samples (12 ms), ω_0 = 86 Hz

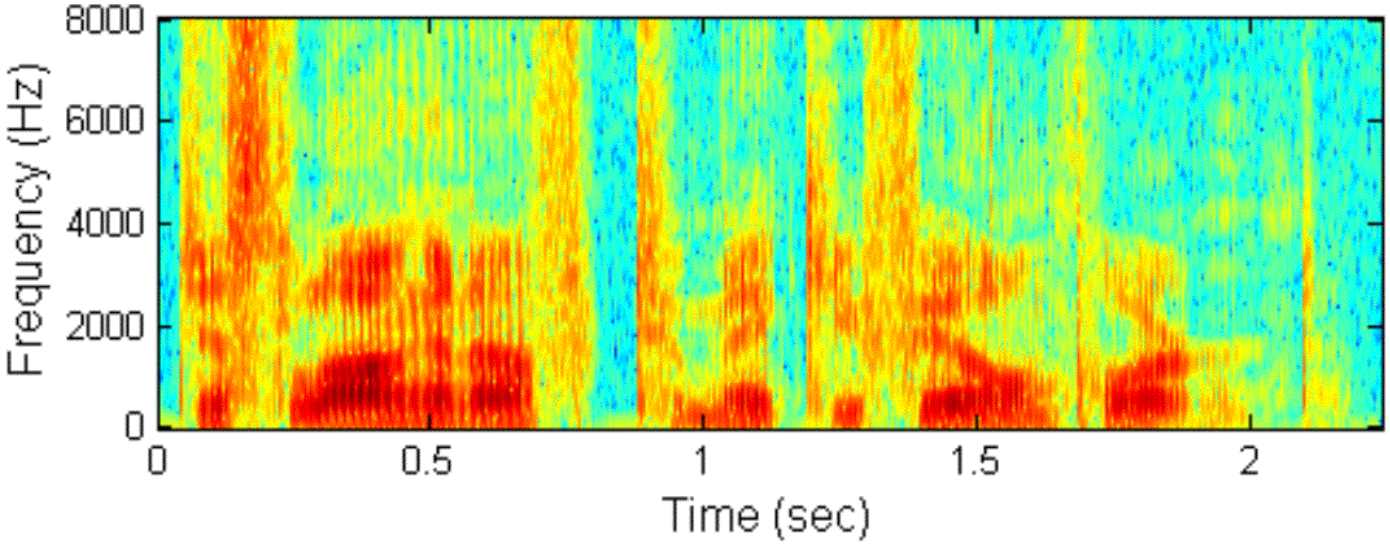
T = 2048 samples (48 ms), ω_0 = 21 Hz

You cannot have high precision of both frequency and time.



Narrowband

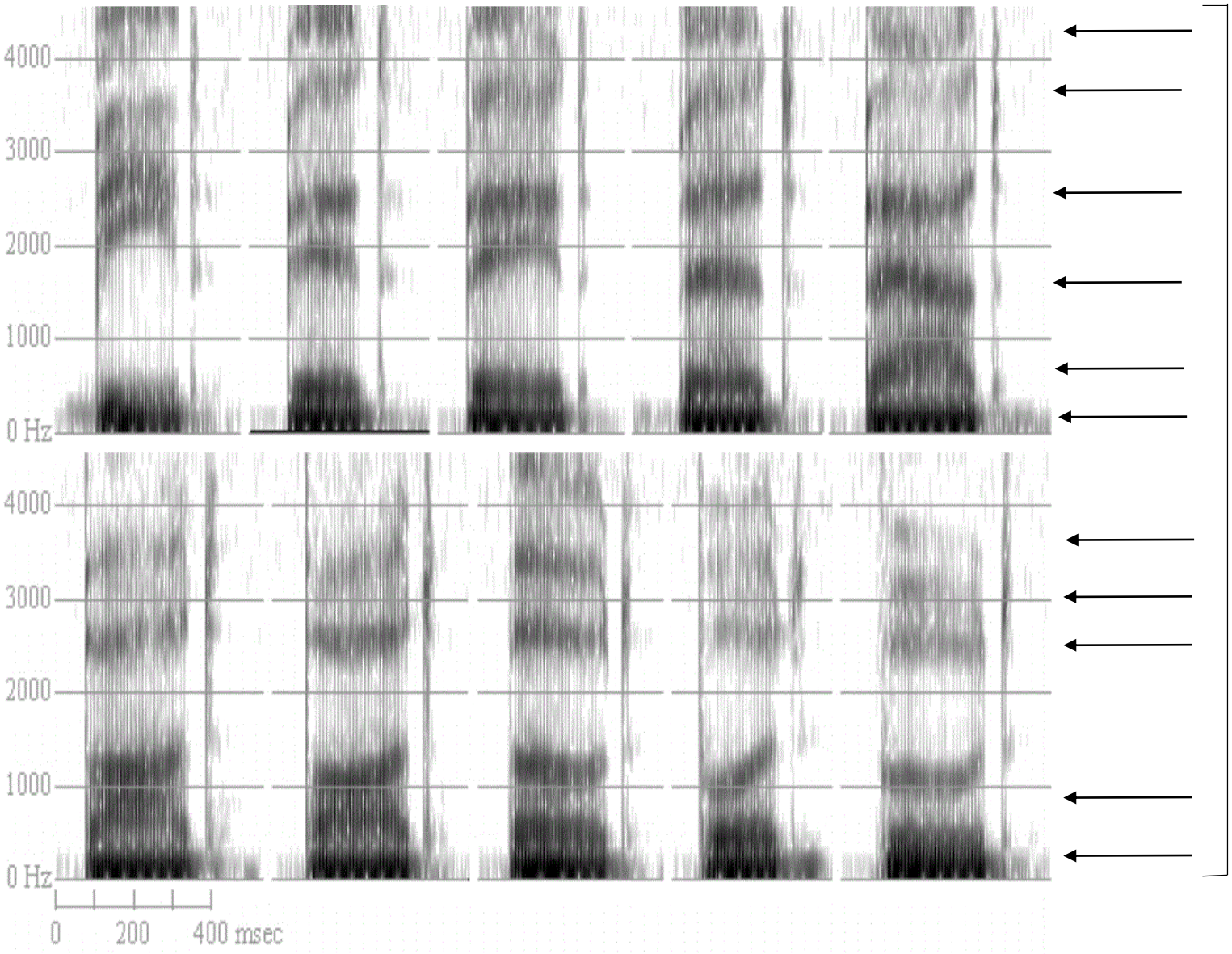
(good frequency resolution, poor temporal resolution … ~48ms)



Wideband

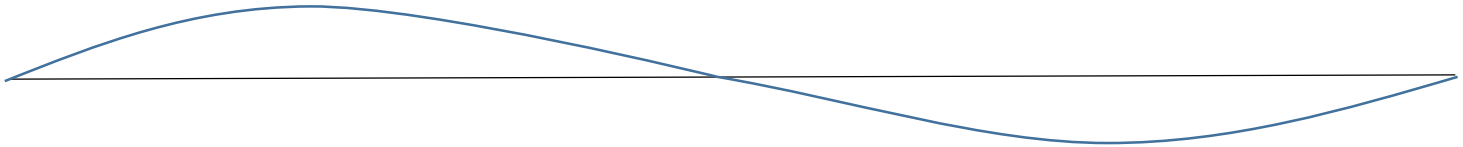
(poor frequency resolution, good temporal resolution … ~12 ms)

Example: Wideband spectrograms of 10 vowel sounds



formants

Spectrogram time scales capture auditory events in the world (e.g. parts of speech, impacts, …) at relatively large time scales.



e.g. period of 12 ms, ω_0 = 86 Hz, $\lambda \sim 4$ meters

These low frequencies play little role in spatial hearing (last lecture).

What are the impulse response functions of auditory filters?

(durations, bandwidths and center frequencies)

Auditory filters

- head related impulse response
- basilar membrane

  http://www.neurosci.info/courses/systems/Nobels/1961%20von%20Bekesy/bekesy-lecture.pdf
- hair cells and ganglion cells in cochlea
- brainstem e.g. MSO, LSO
- cortex A1 (later today … larger time scales)

Auditory filters

Classical experiments used pure tones and/or noise.

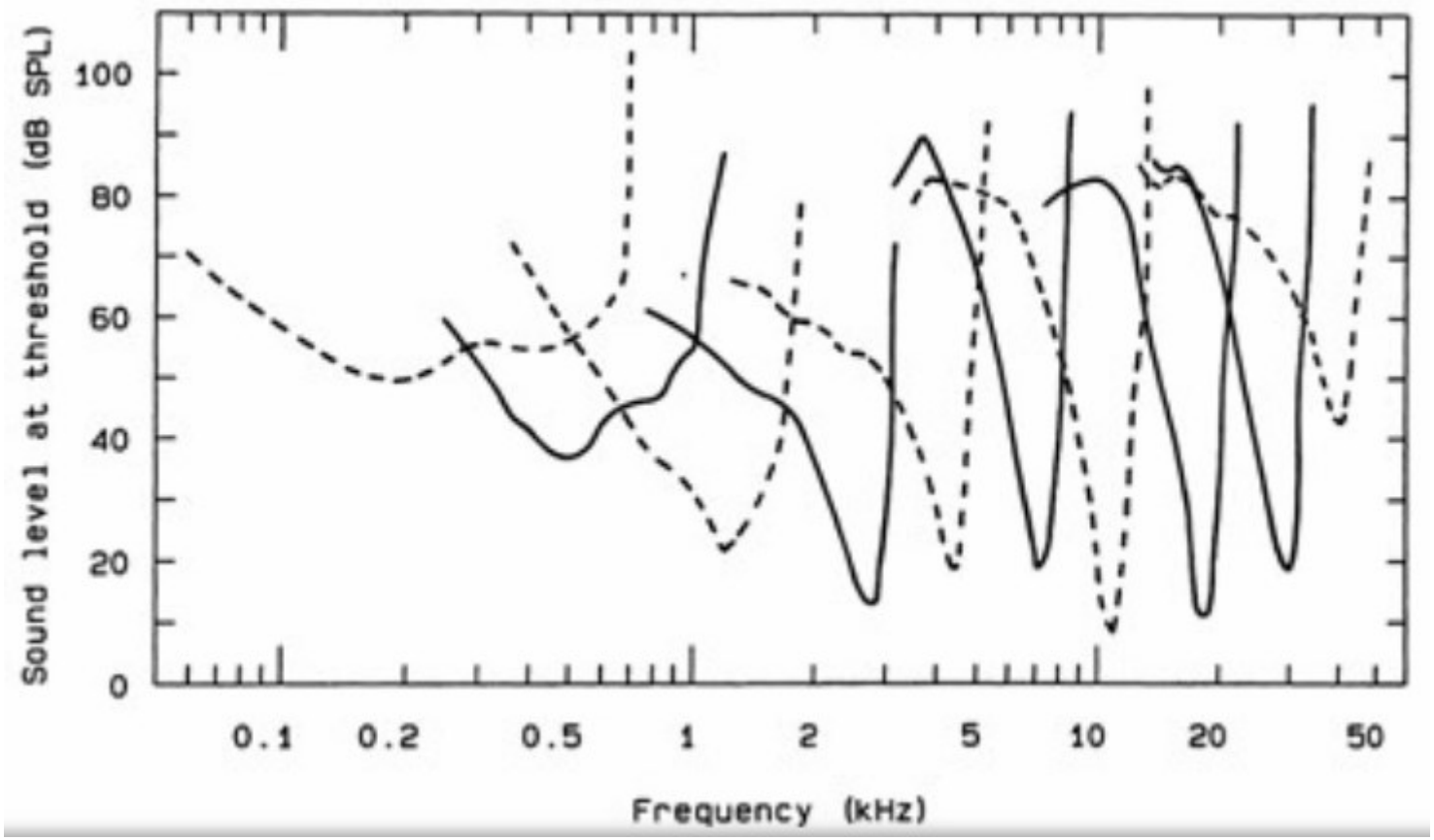
(starting in 1940's and going for 50 years)

- recording from single cells

  (BM, nerve fibres in cochear nerve, brainstem)

- psychophysics e.g. masking

Example: Frequency tuning curves (thresholds) for different ganglion cells to pure tone stimuli

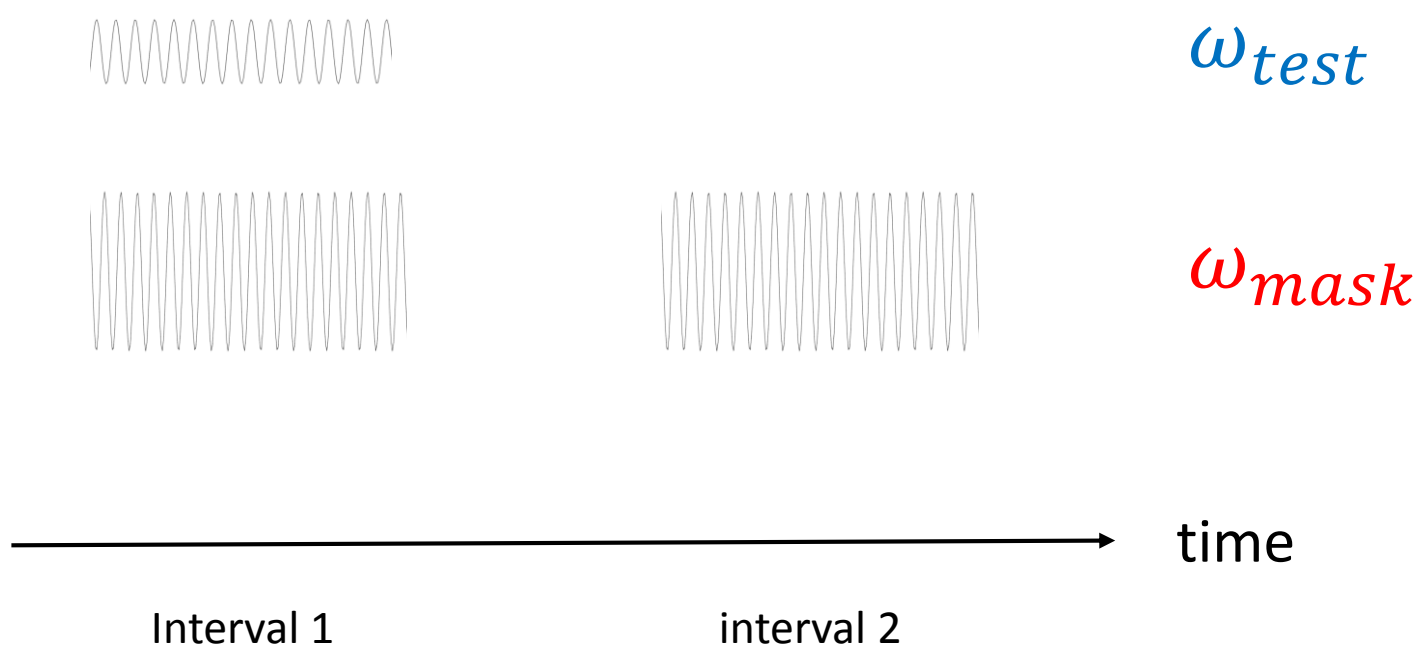


Psychophysical Masking

How does presence of one frequency component affect our ability to hear other frequency components?

Two similar frequencies mask each other more than two different frequencies.

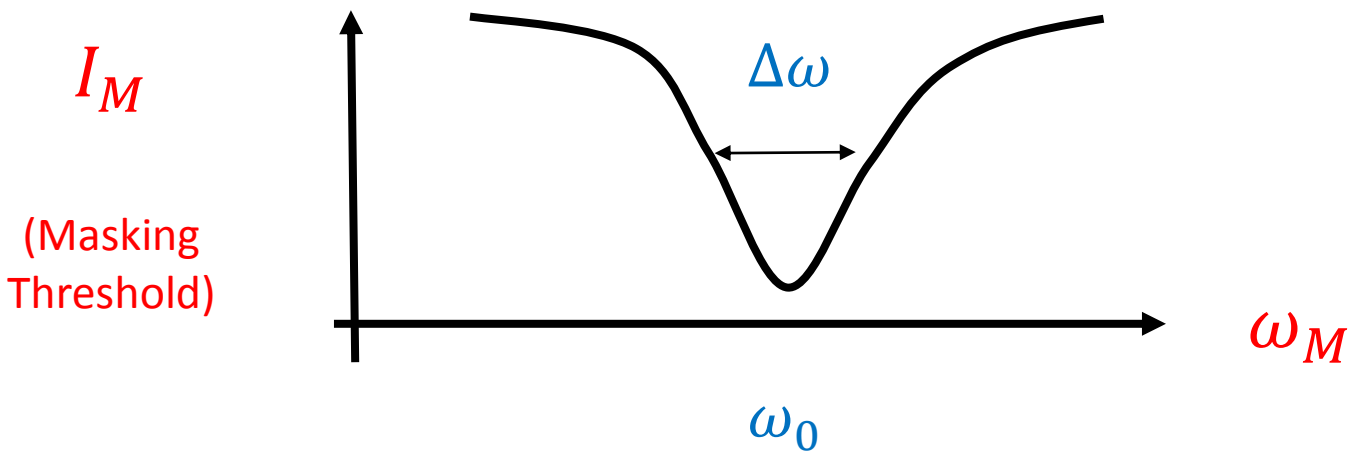
Example Masking Experiment



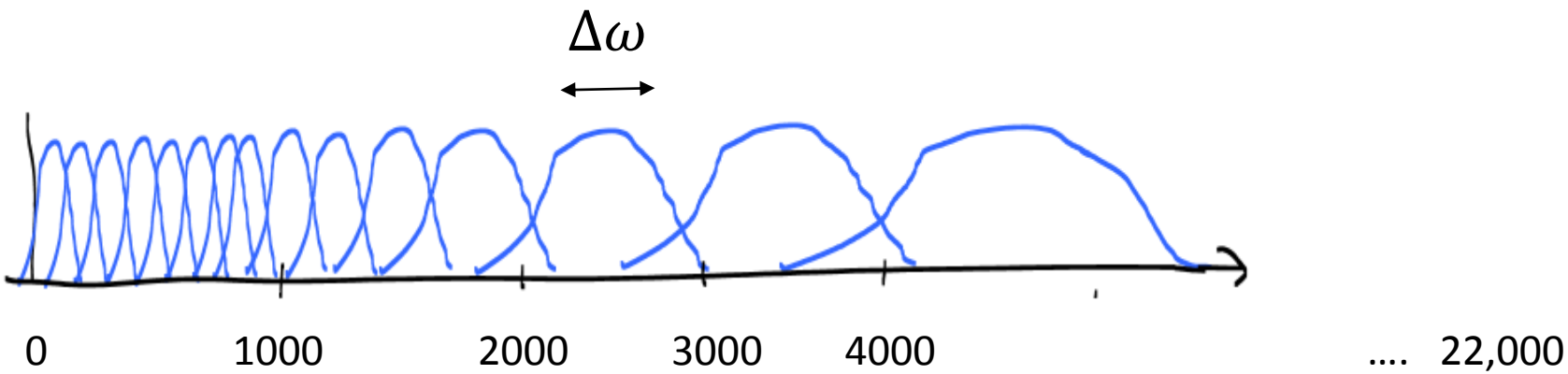
Task: Which interval contains the test tone?

For each test frequency ω_0 with some given SPL,
For each masking frequency ω_M
Measure a masking threshold $I_M(\omega_M)$

Define "*critical bandwidth*" for ω_0 by $\Delta\omega$.



Auditory filters: typical bandwidth model



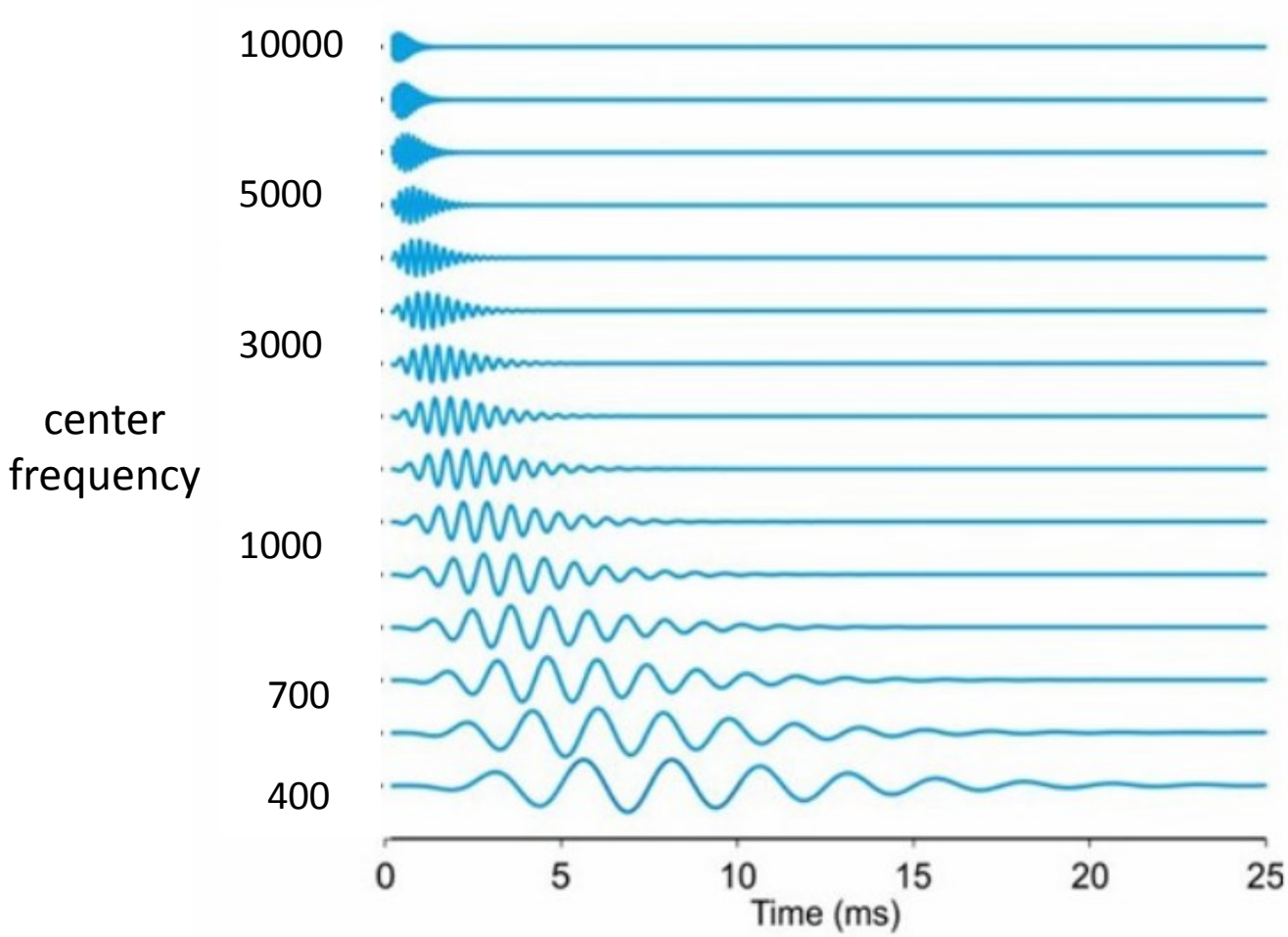
$\Delta\omega$ is ~100 Hz for center frequency up to 1000 Hz.

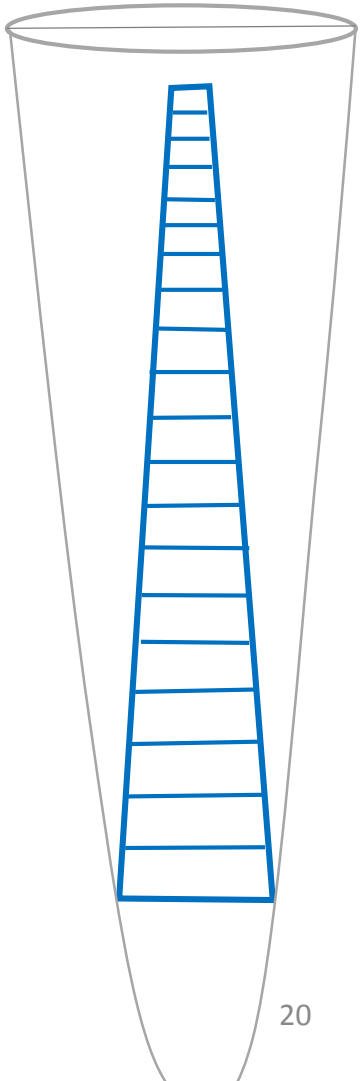
$\Delta\omega$ is ~ 1/3 octave from 1000 Hz up to 22, 000 Hz.

Gammatone filter model

Similar to Gabor filters but window is asymmetric.

(Also, note shifted in time to enforce *causality*.)

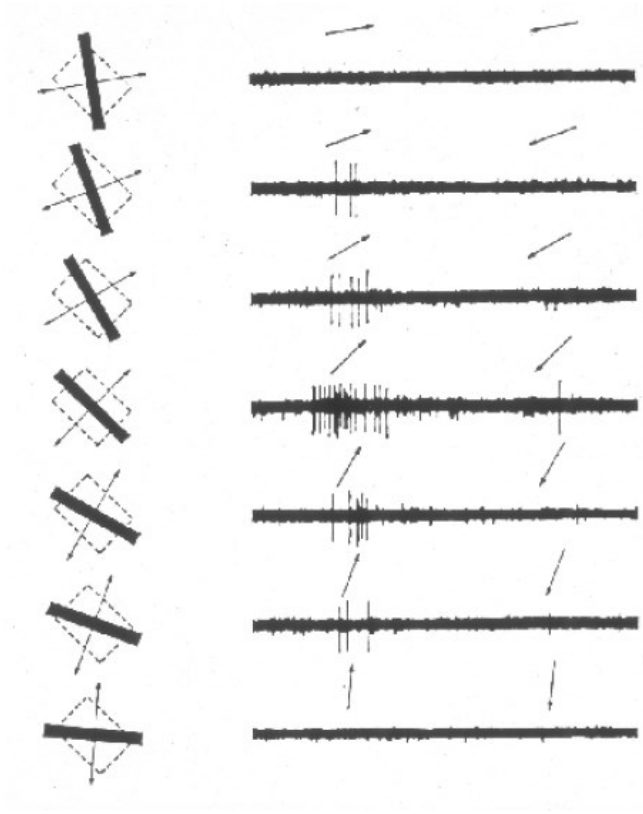




Auditory filters

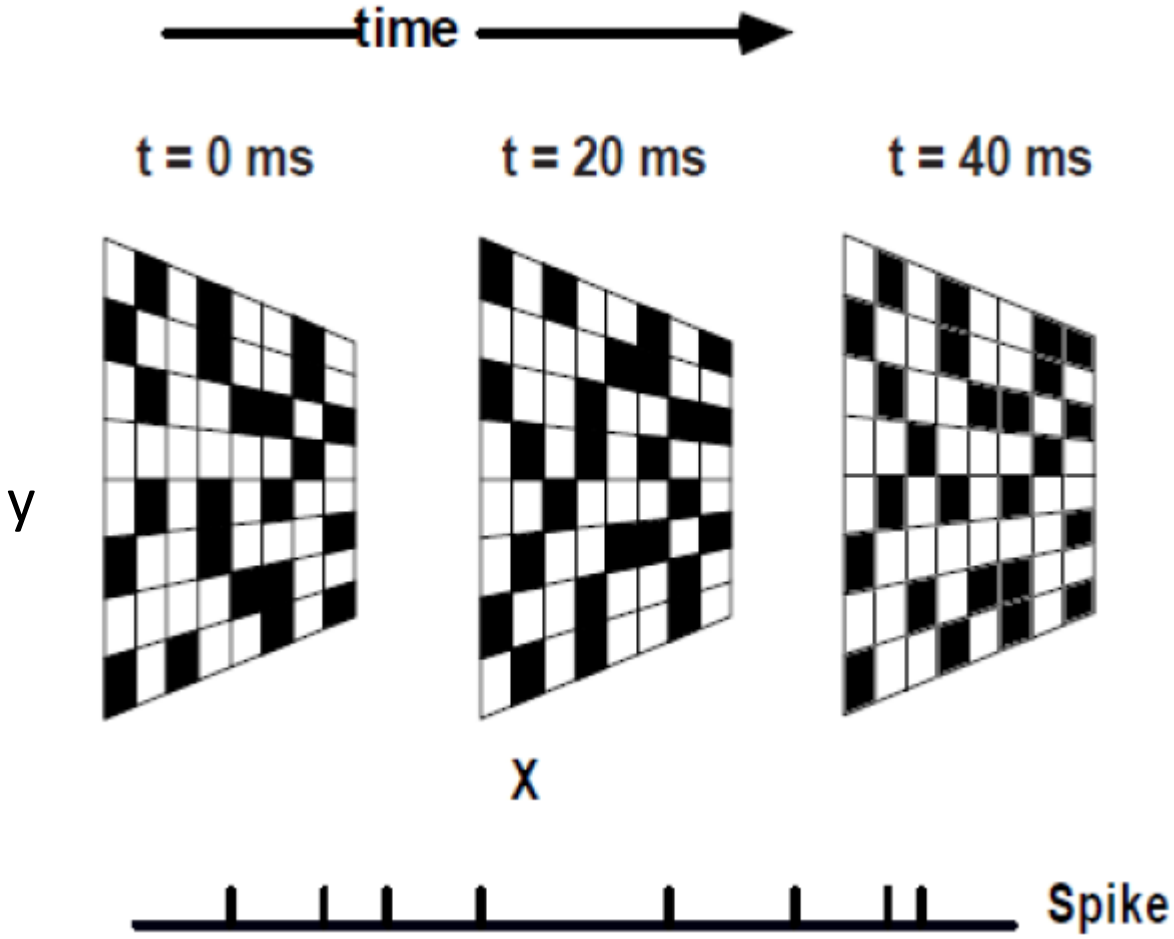
- head related impulse response
- basilar membrane
- hair cells and ganglion cells in cochlea
- brainstem e.g. MSO, LSO
- cortex (A1 and beyond)

V1: recall Hubel and Wiesel (1962)



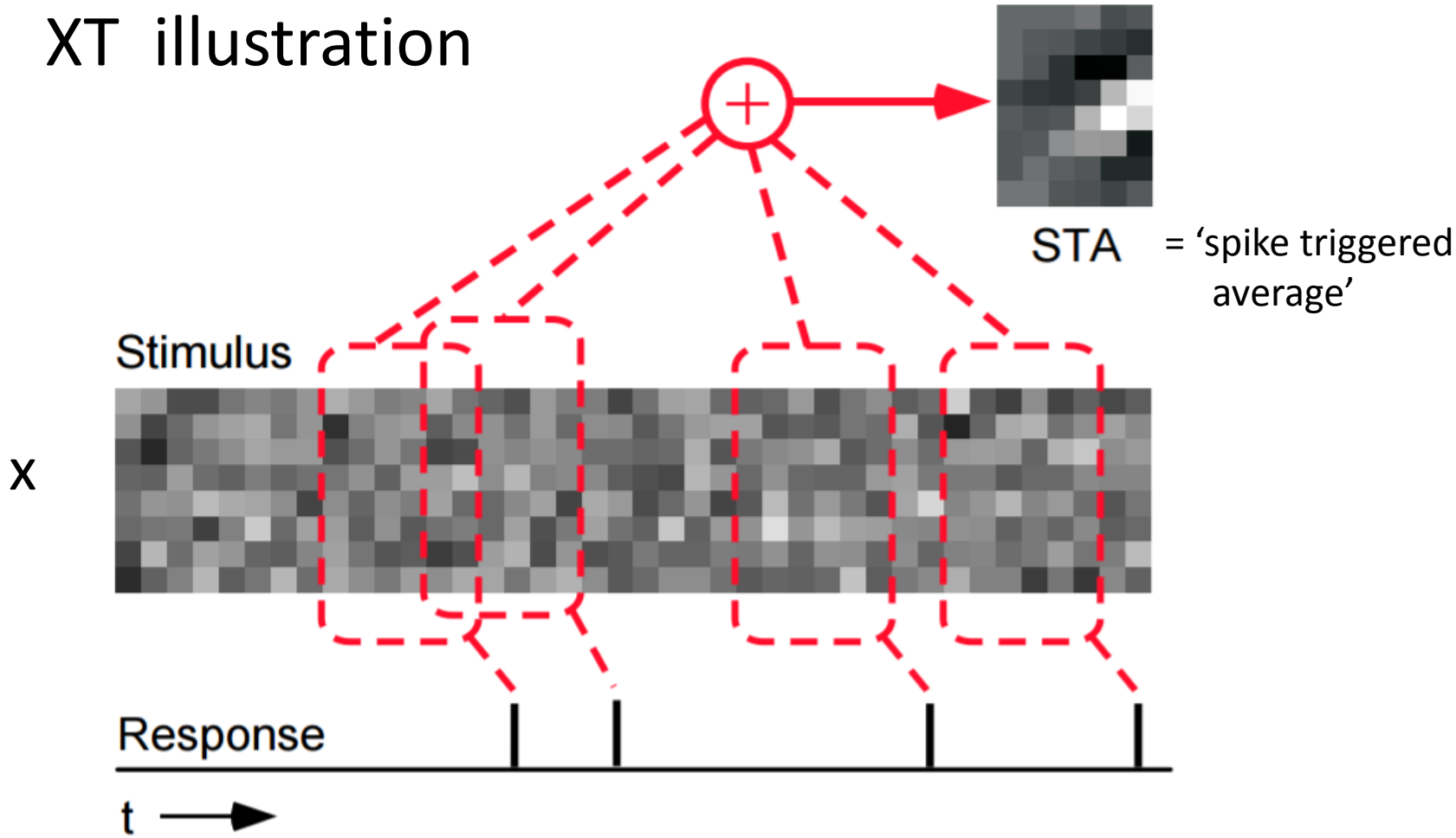
Such a stimulus works well if you *already know* the cell is orientation and motion selective.

Q: What to do if you don't know anything about the receptive field?
A: Compute "spike triggered average".



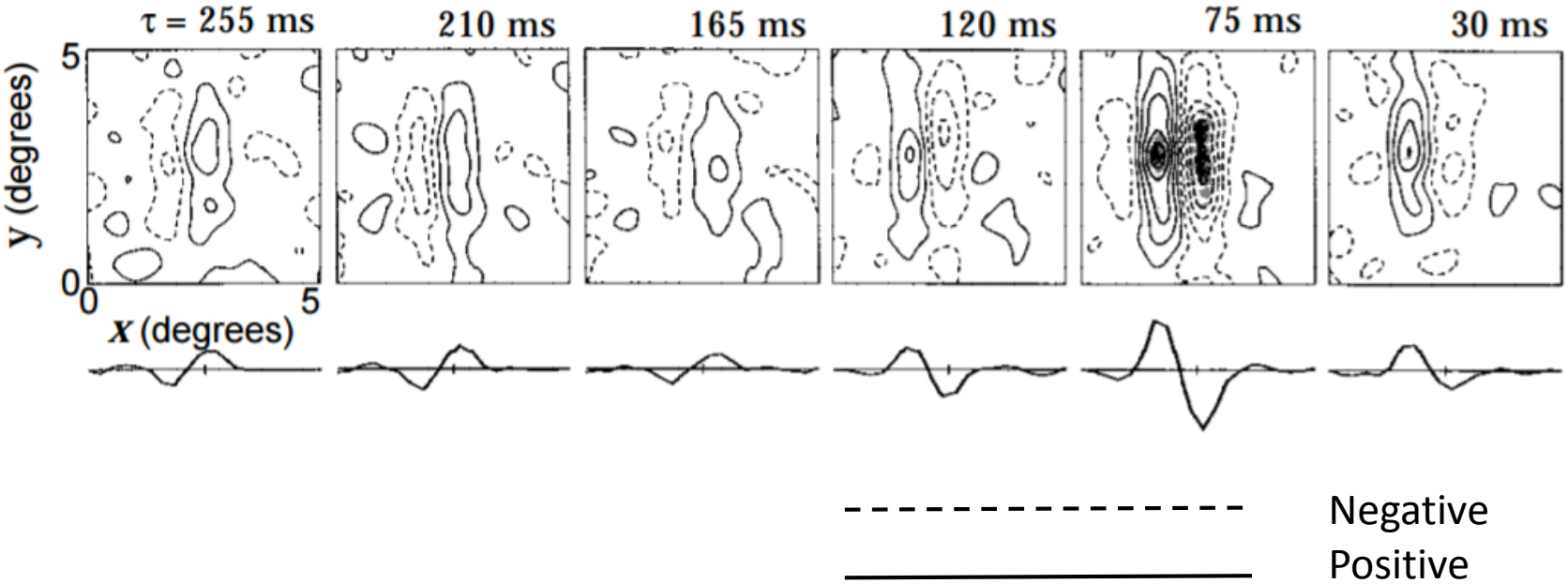
Use random input (often white noise). *What is the average spatio-temporal stimulus that preceded the spikes?*

e.g. XT illustration



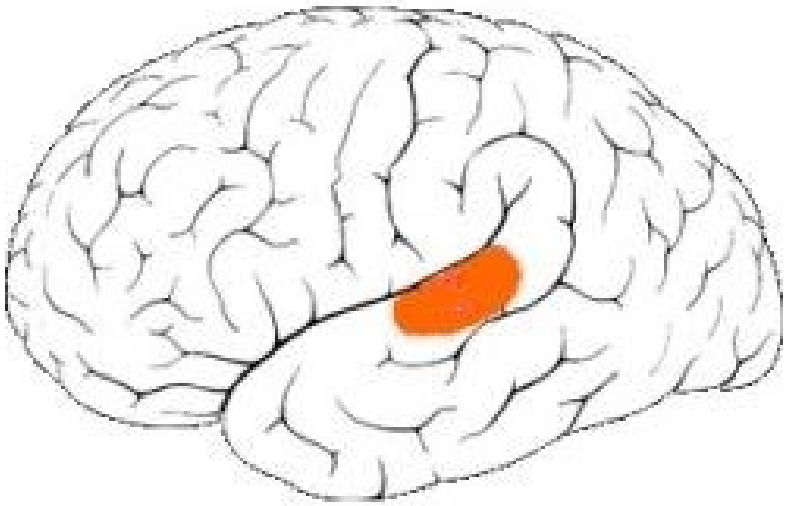
Real data for V1 receptive field (XYT)

Spike triggered average stimulus (backwards in time). Spike at t=0.



[DeAngeles 1995]

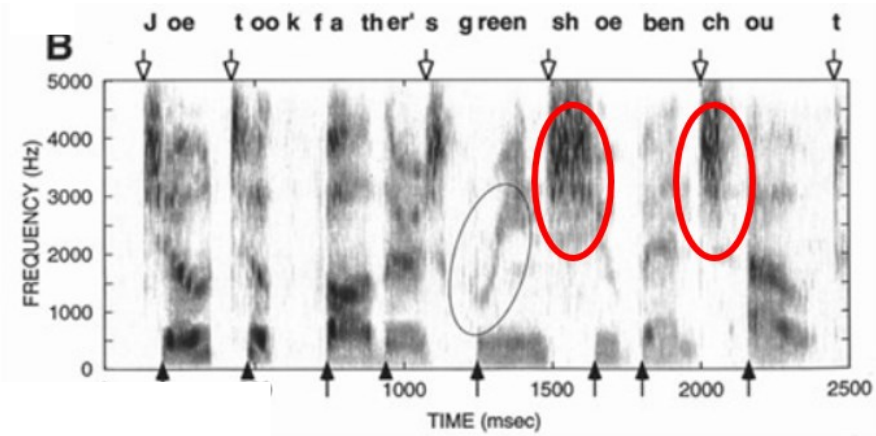
Auditory Cortex Receptive Fields



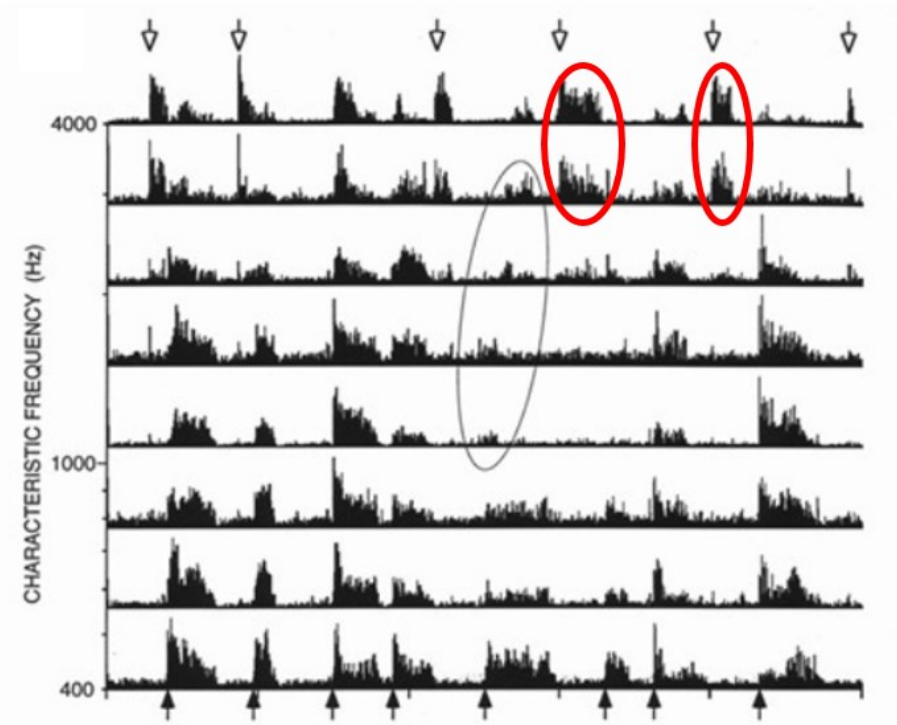
Inputs to A1 and have been spectrally bandpass filtered.

There is ~ no more phase locking to stimulus sound.

Example of responses of 8 auditory nerve fibres to a voice sound



Spectrogram of voice saying "Joe took father's green shoe bench out".



Spike histograms of *auditory nerve* fibres (cat) with different peak ("characteristic") frequency sensitivities.

[Delgotte 1997]

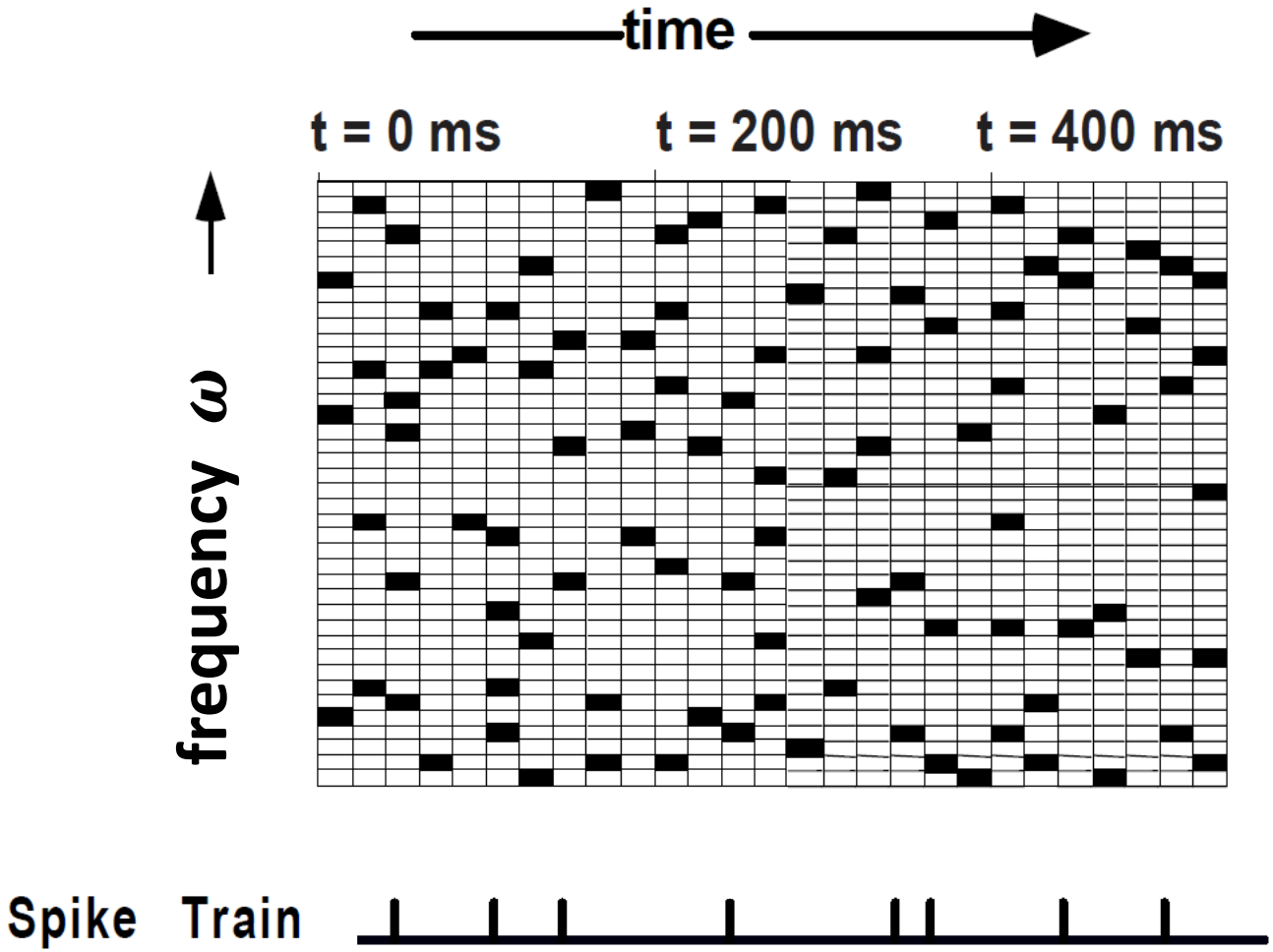
What stimuli to use? (Cats don't understand human speech, so it unlikely we would find cells tuned for it.)

Recall Hubel and Wiesel had first tried using center-surround stimuli for cells in V1.

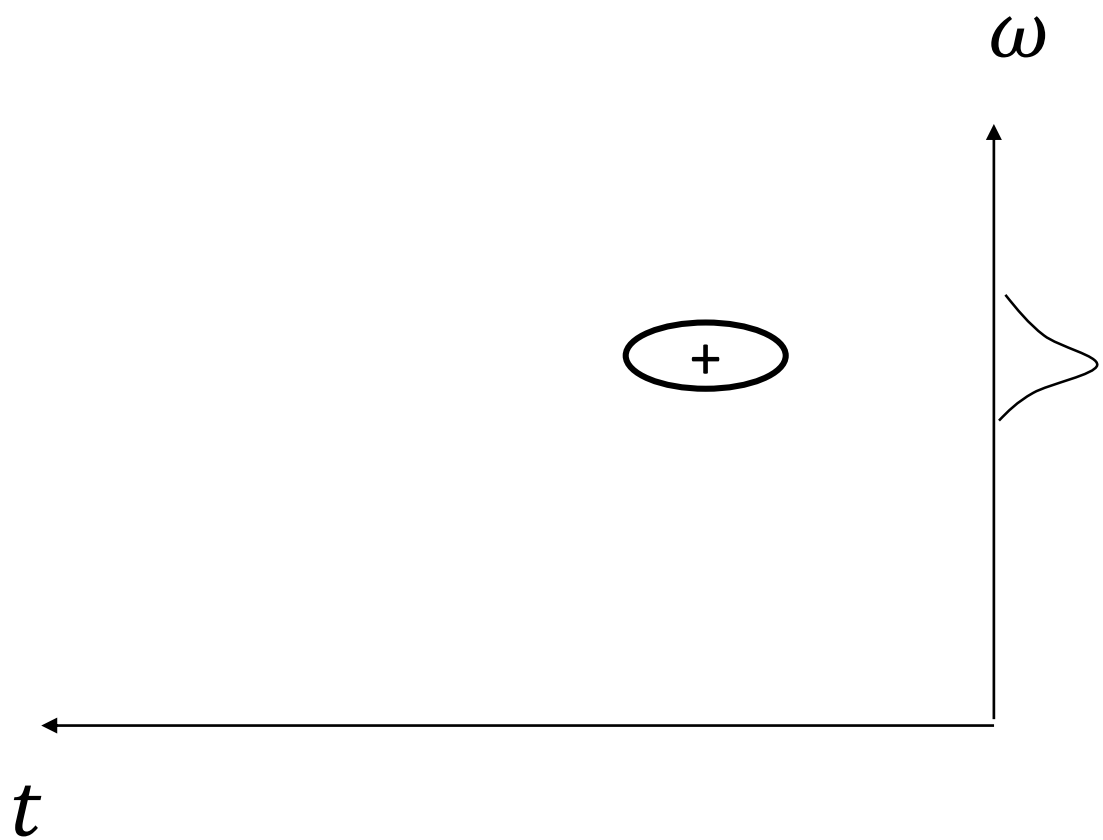
The analogy in audition would be to use the same bandpass stimuli used for auditory fibres.

Any other ideas?

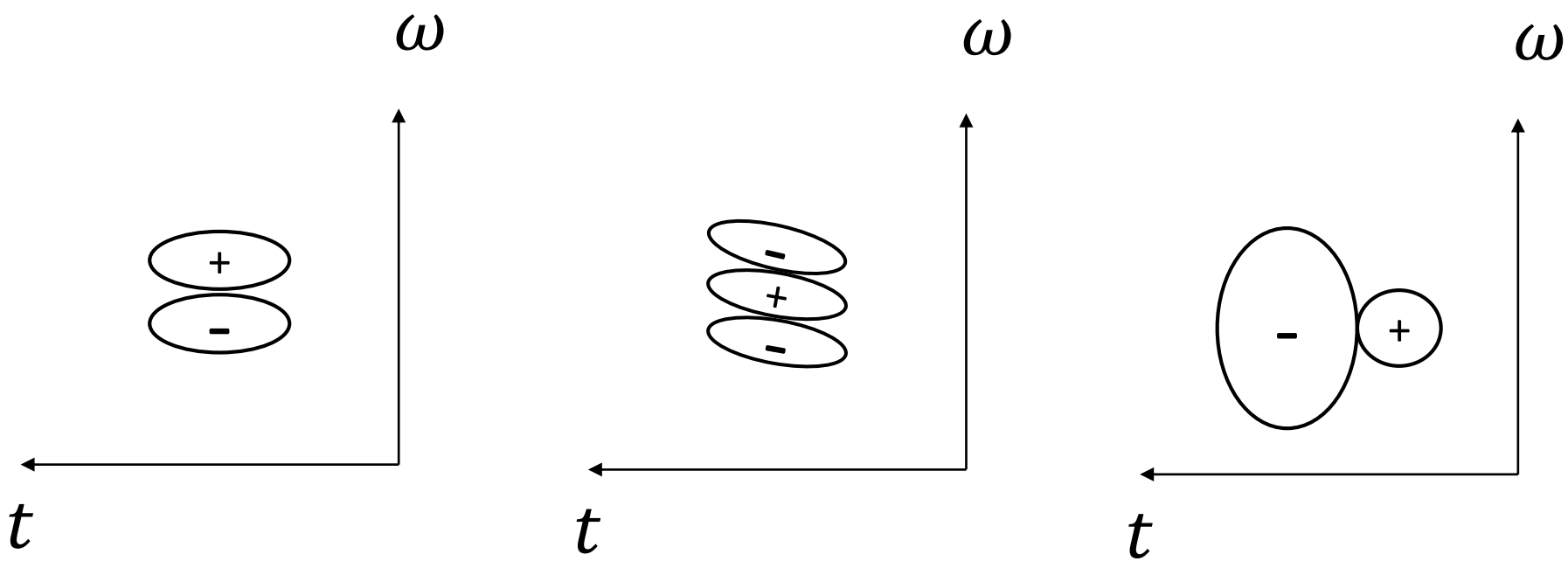
Random “chord” stimuli [deCharms, 1998]



What spike triggered average should we expect from a bandpass cell ?

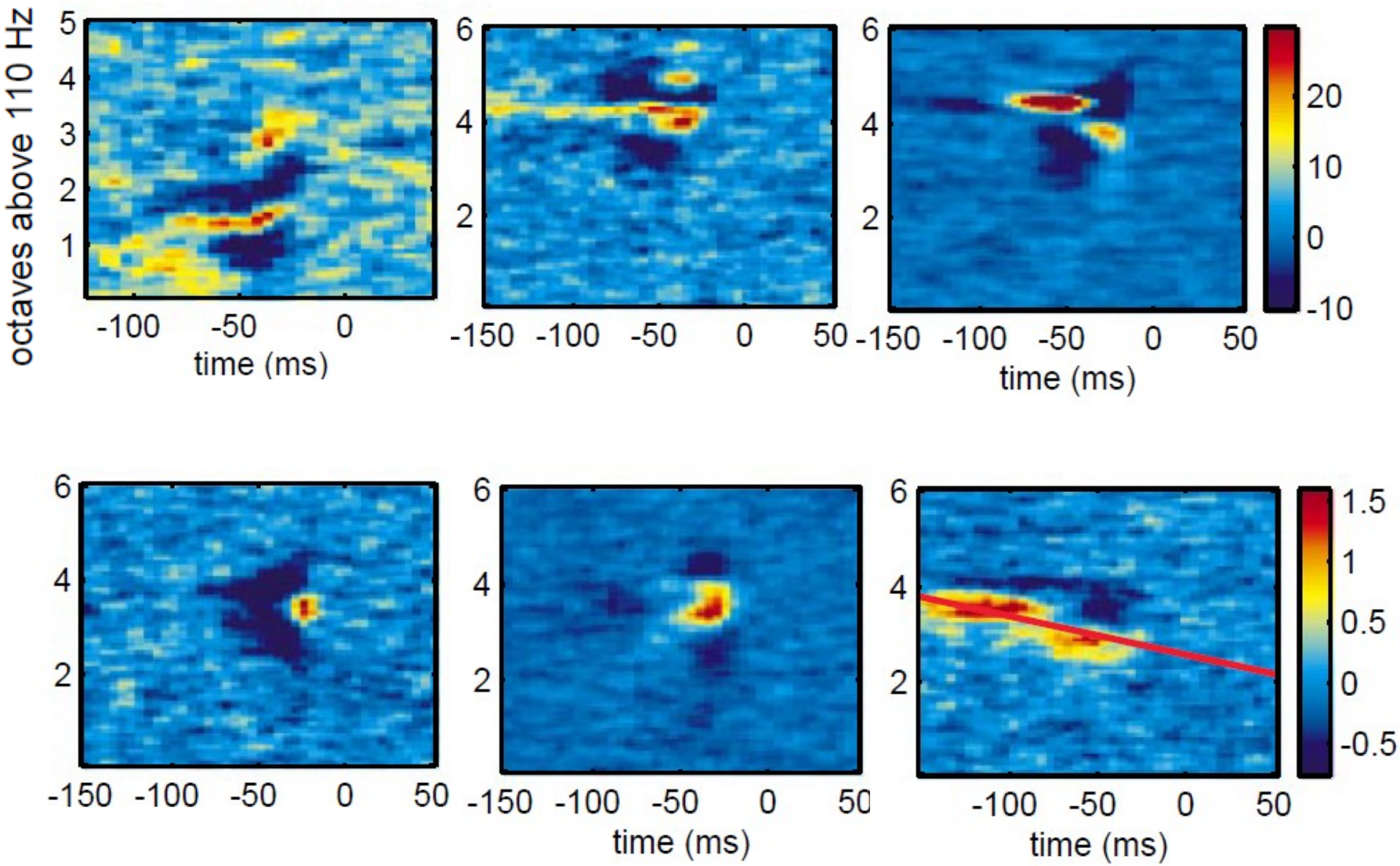


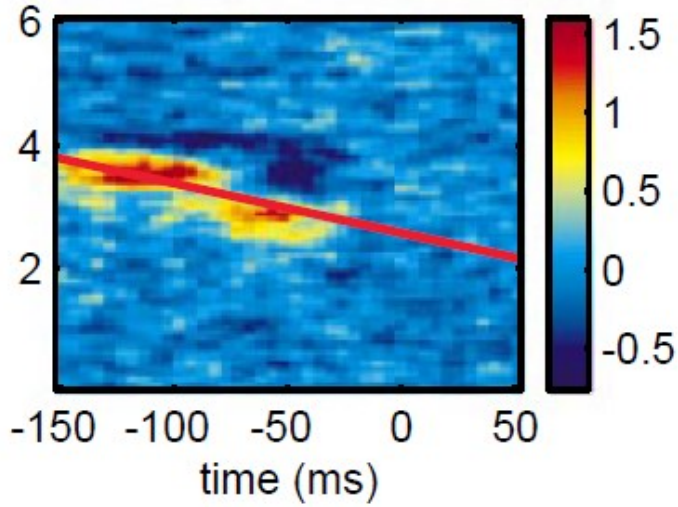
Do we find more interesting cells such as… ?



Examples: Spectro-temporal receptive fields of A1 neurons

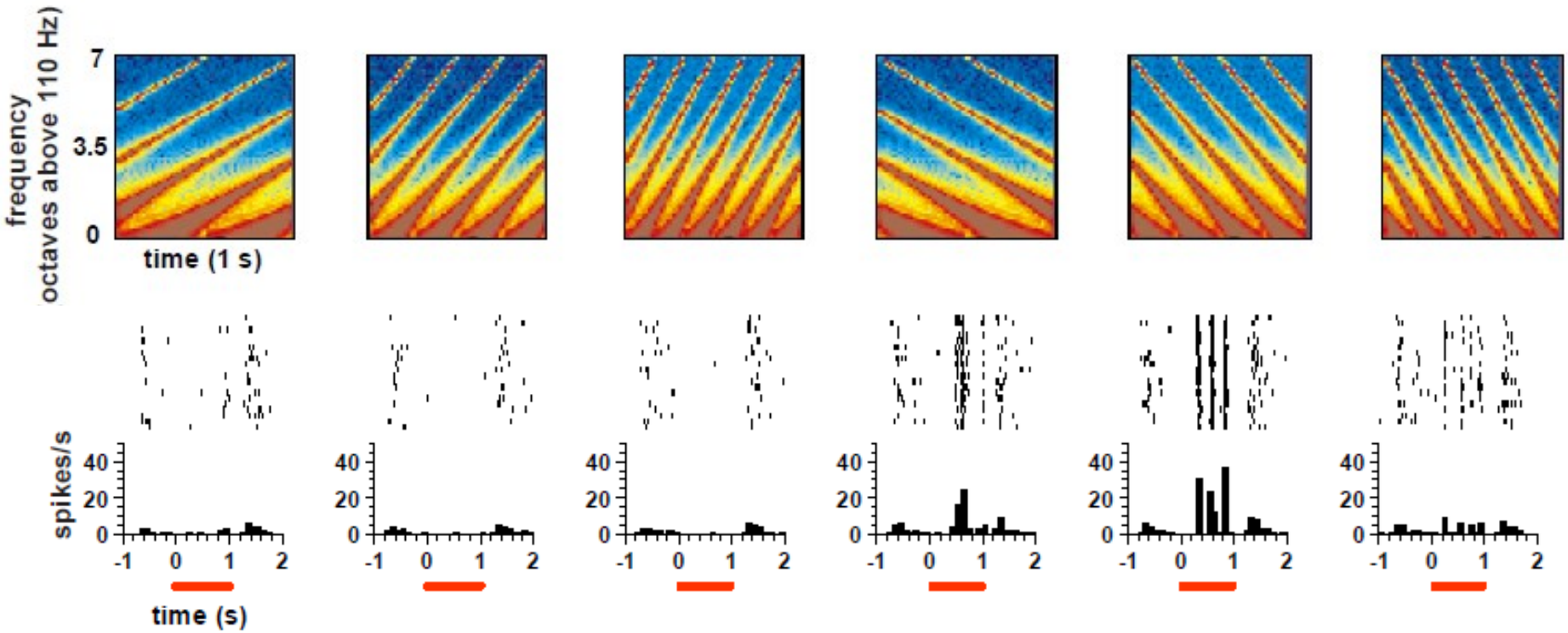
[de Charms, 1998]





Orientation (ω, t) selective ?

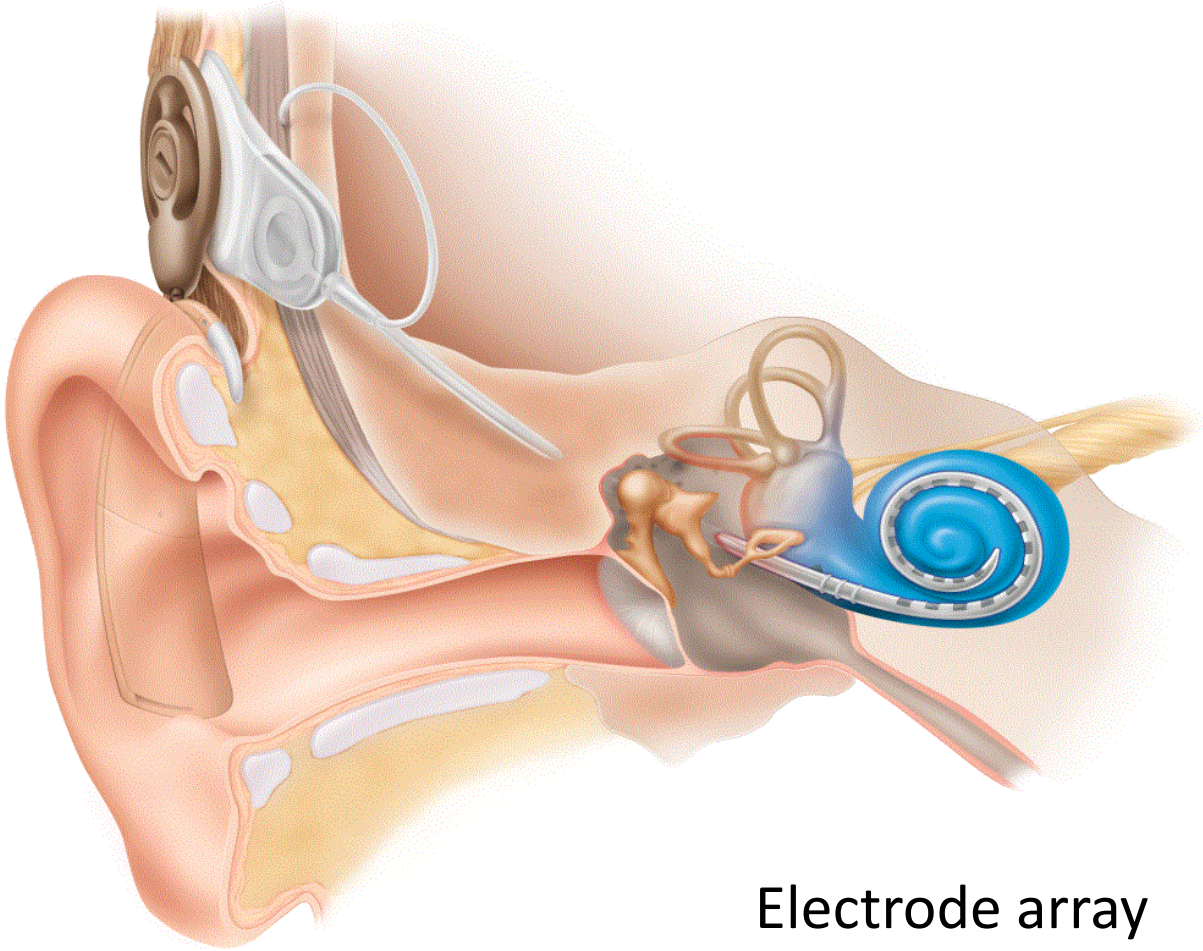
Verify the responses of the above cell to a tone and its harmonics, changing over time:



ASIDE: Two Applications

Cochlear implants are used for profoundly deaf people whose hair cells destroyed by disease but auditory nerve is intact.

Microphone + speech/sound processor



Electrode array (inserted into cochlea)

MP3: Data Compression

Simultaneous masking: what I mentioned earlier

Forward masking:

Sound at time t can mask sound at time $t + \Delta t$ and nearby frequency bands, even if Δt is greater than auditory (gammatone) filter.

In both cases, you can use fewer bits to code sound and listeners won't notice.