Cortical Representation

CMSC 828D / Spring 2006

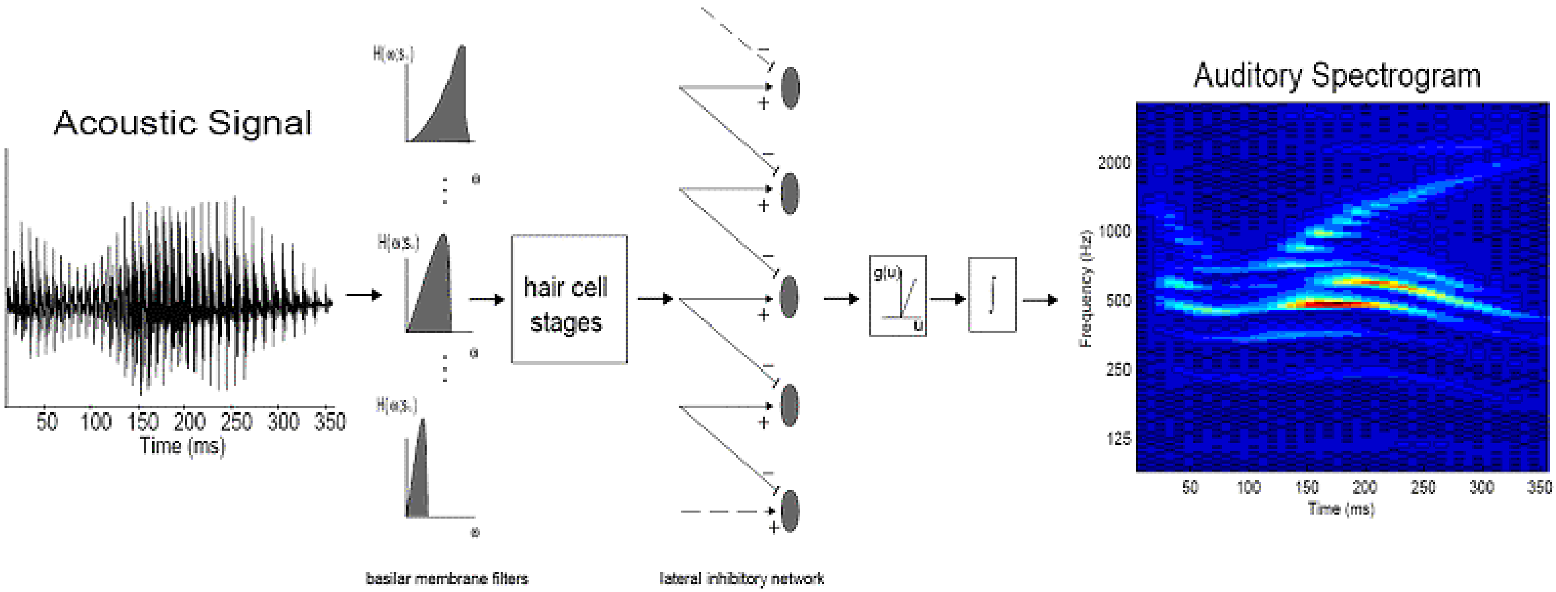
Lecture 24

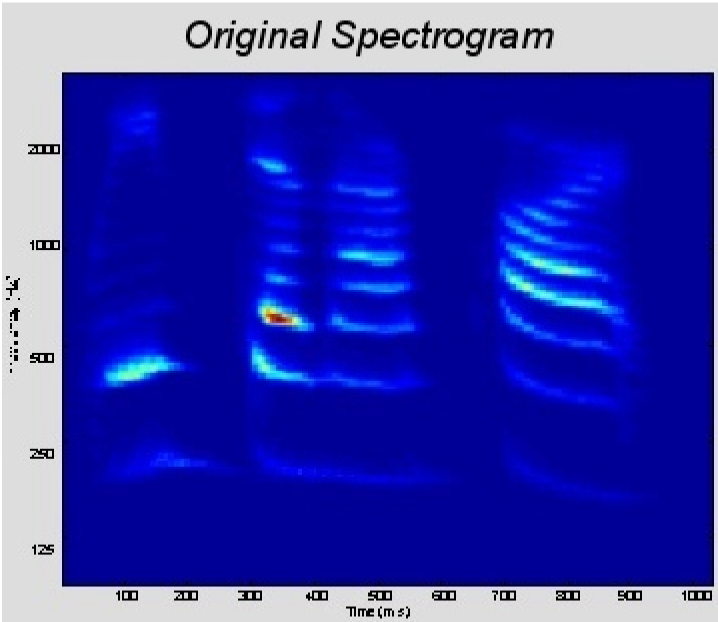
(some slides are adapted from
Dr. S. A. Shamma, University of Maryland)

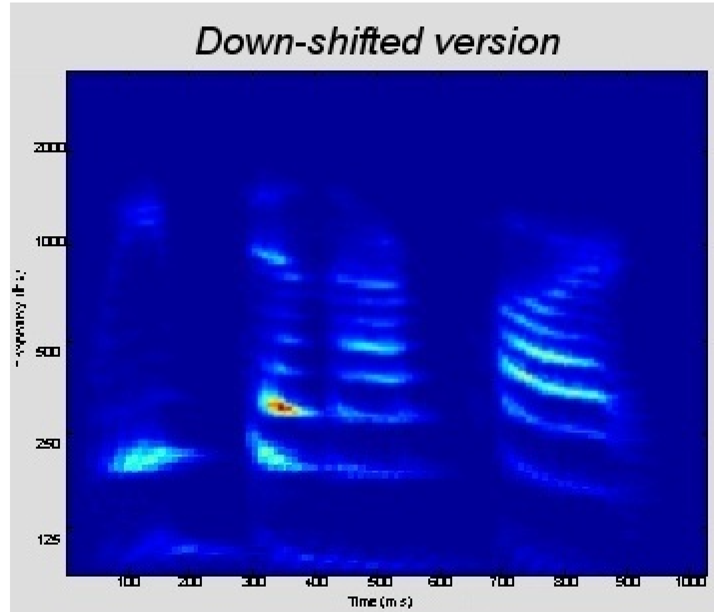
Cochlear Processing

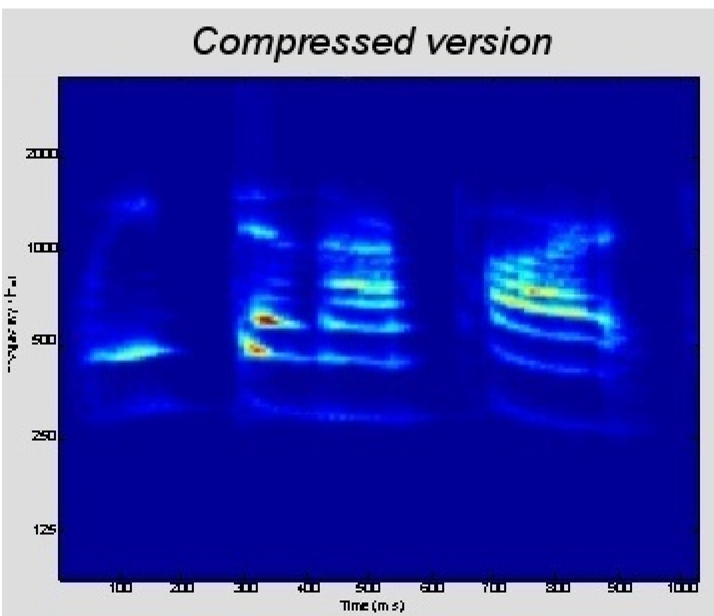
- Recap of Lecture 9
- Cochlea performs frequency analysis
  - Along the basilar membrane, different frequencies resonate at different points
- Result is the auditory spectrogram

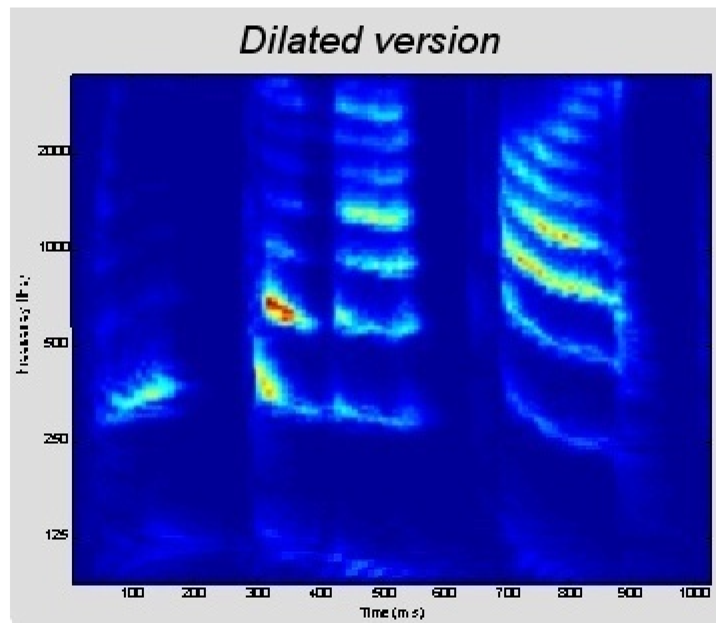
Early Auditory Stage




Original Spectrogram
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (m s)


Down-shifted version
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (m s)


Compressed version
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (m s)

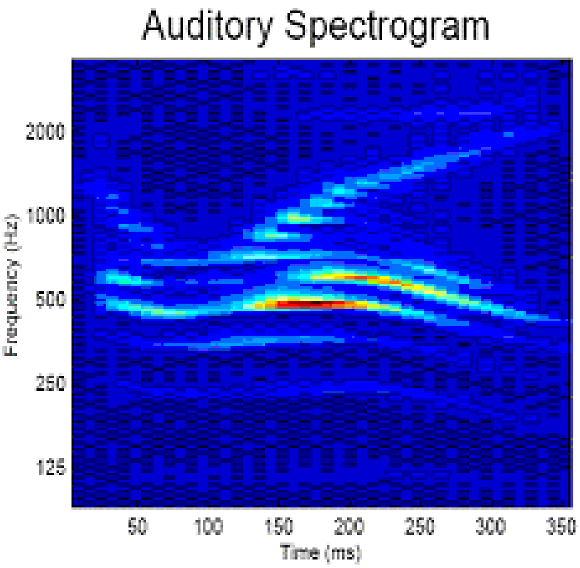

Dilated version
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (m s)

Central Auditory Stage

- Recorded response of neurons in the brain
  - Auditory cortex of ferrets
- Selective response to particular patterns in auditory spectrogram
  - It is hypothesized that these neurons pick up “features” used in sound recognition

Decomposition Basis

- Look at auditory spectrogram
- Groups of frequencies sweeping up or down
  - Characterized by:
  - Spacing in frequency
  - Rate of frequency change

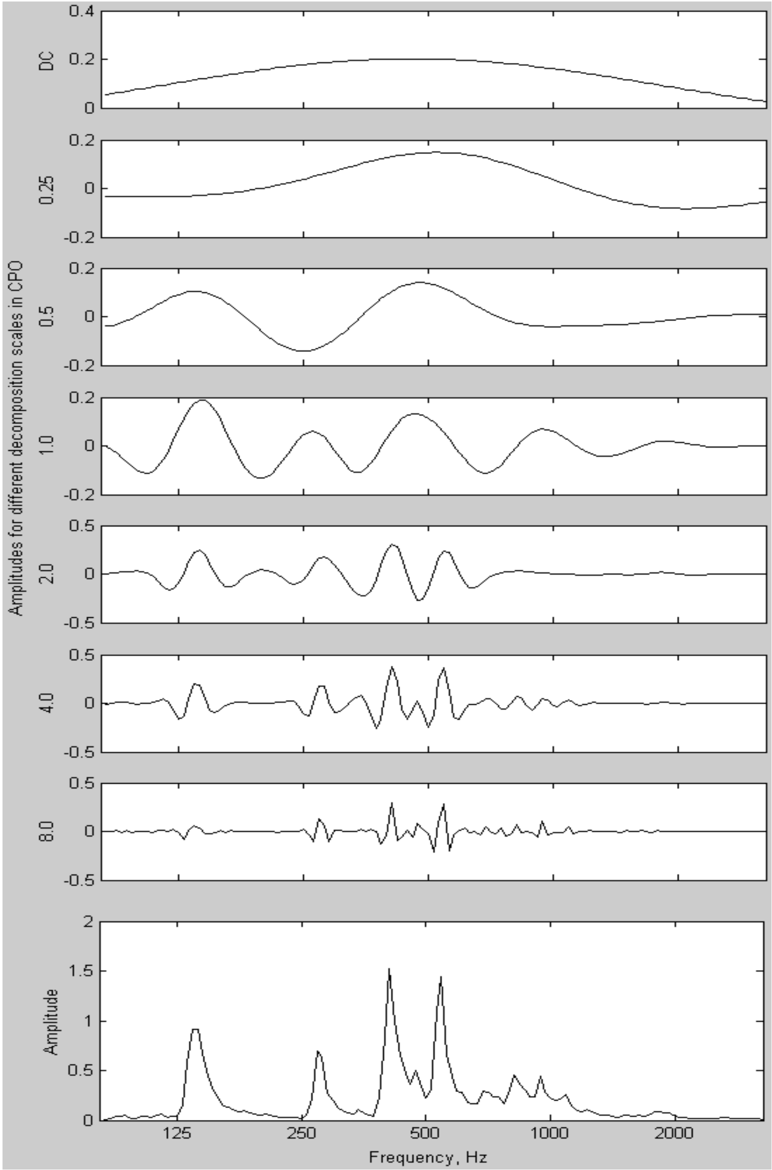


Decomposition Basis

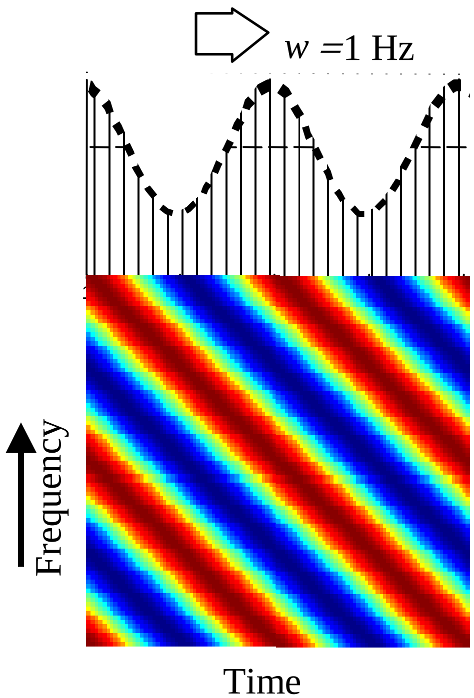
- Sound ripple is a sound that has a group of frequencies
  - A given interfrequency spacing (called “scale” and measured in cycles per octave, CPO)
  - A given rate of frequency increase/decrease (called “rate” and measured in Hz)

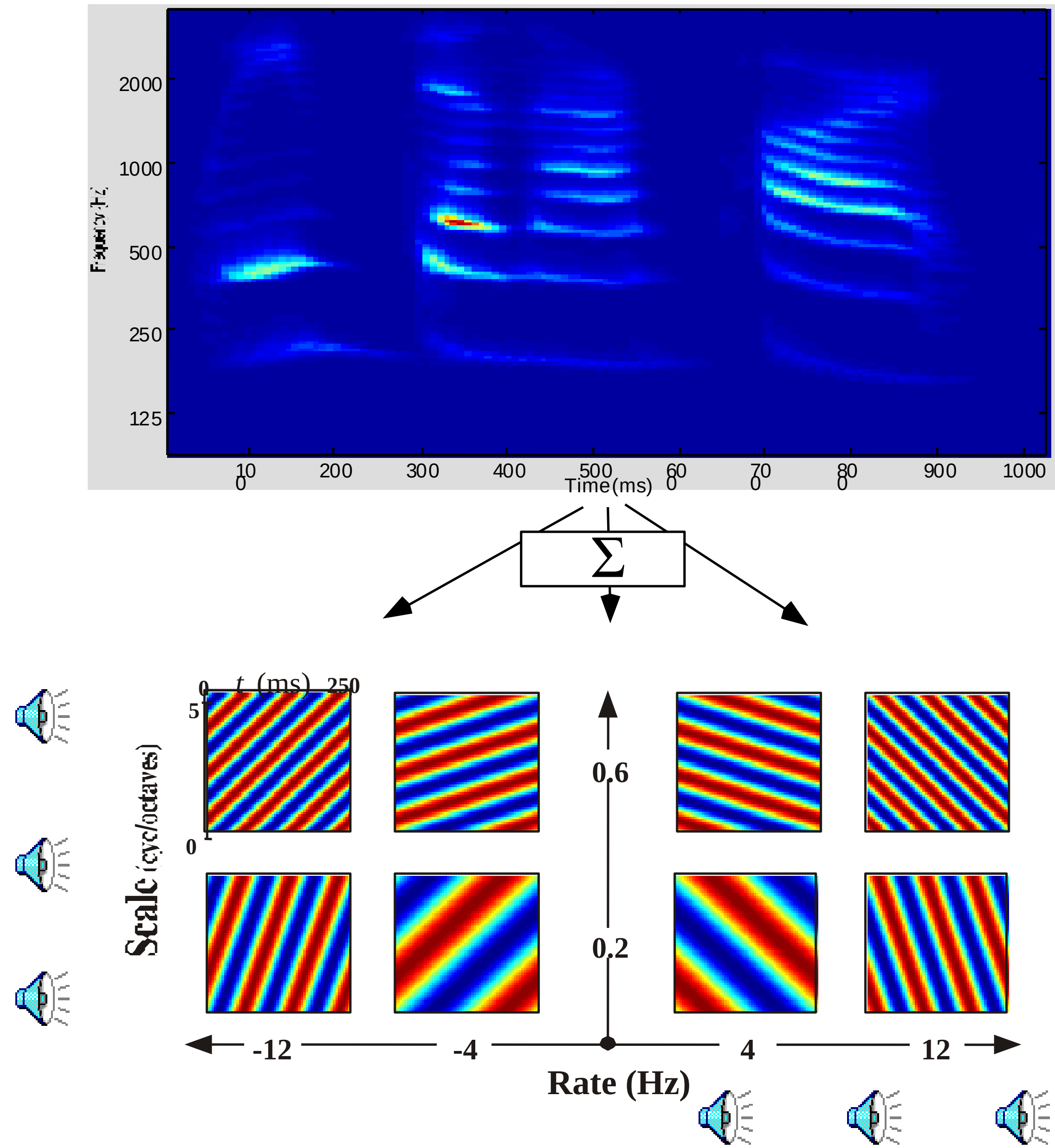
Scale-only Decomposition

- Preliminary example
- On the bottom, there is a time slice of a spectrogram
- Top plots show decomposition into features of various widths (in CPO)



Rate-scale Decomposition

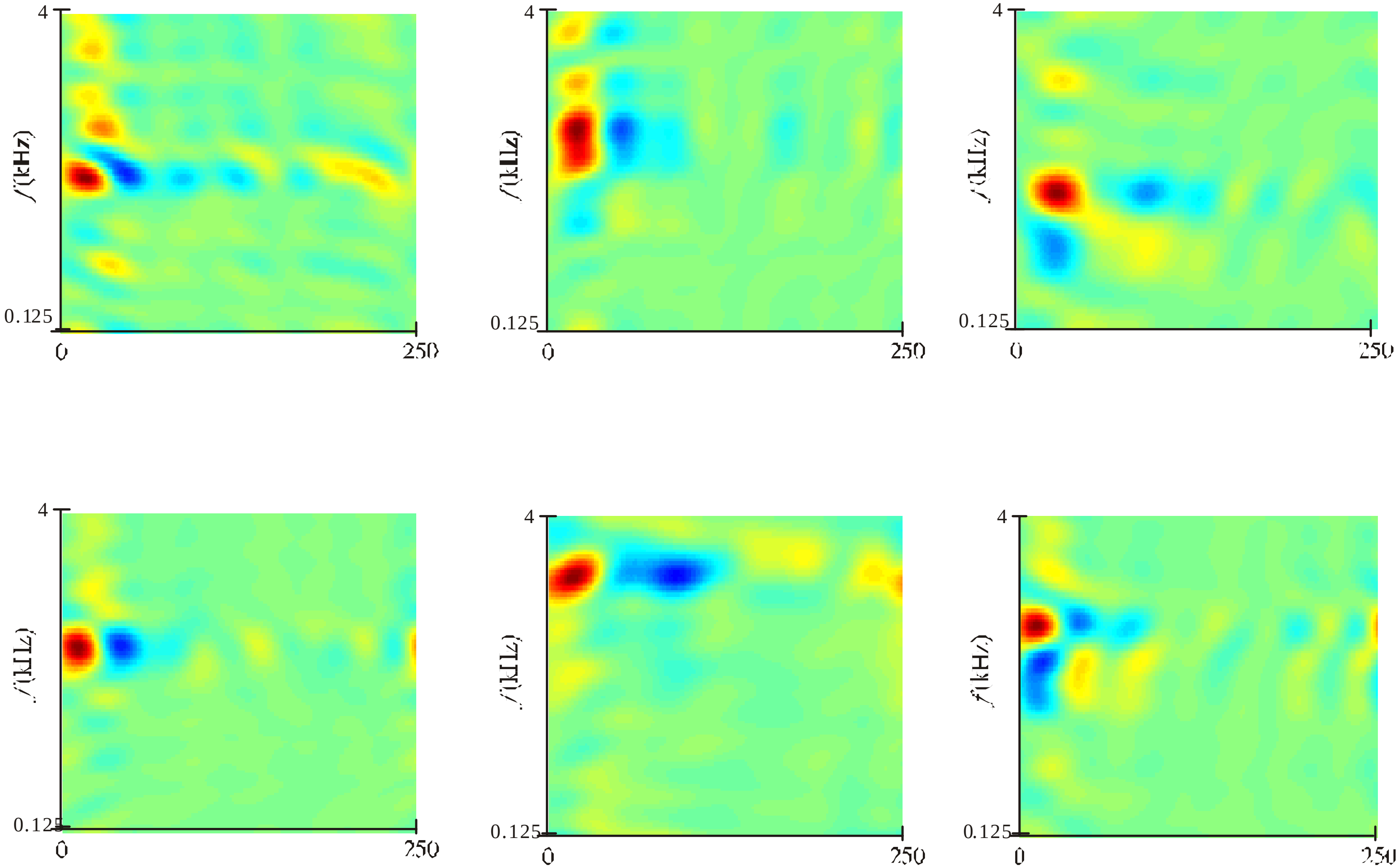




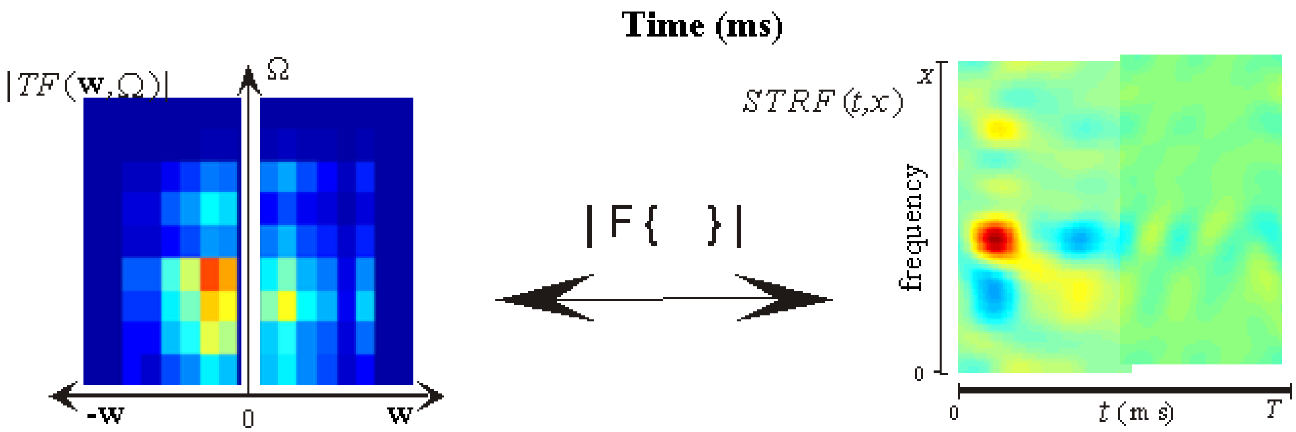
STRF

- Particular neurons respond best to some combination of rate and scale
- Spectro-temporal response field
  – Plot neuron response versus scale and time
  – Rate then is determined from time
- Experimentally collected evidence

Examples of Different STRF Shapes

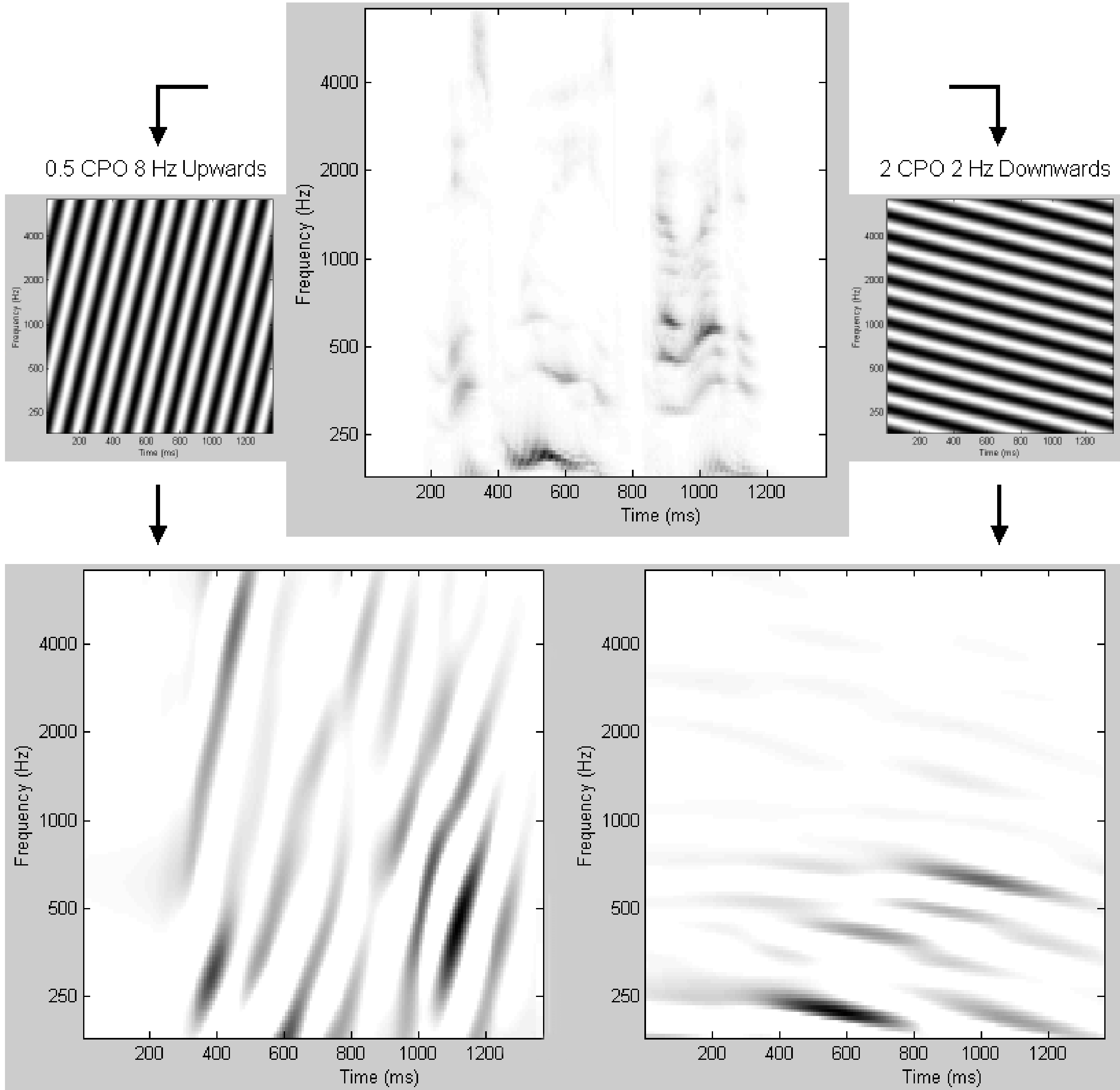


STRF to Scale-Rate Plot



Cortical Decomposition

- Differently-tuned neurons at each frequency
- Neurons are sensitive to:
  - Scale range: 0.125 to 8 CPO
  - Rate range: 2 to 16 Hz
  - Upward and downward moving ripples
- Output of a neuron is high when the input matches the tuning



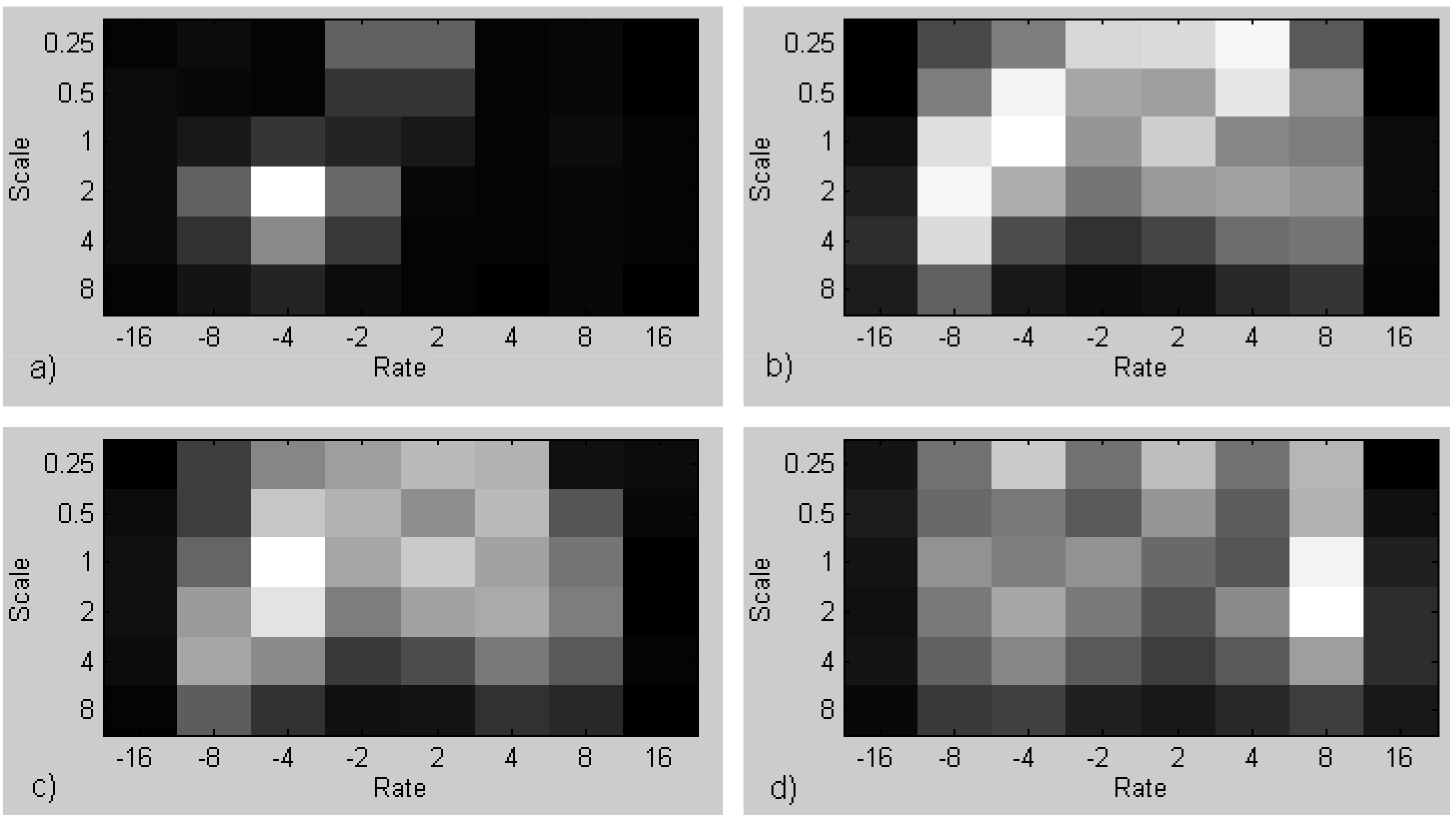
0.5 CPO 8 Hz Upwards
2 CPO 2 Hz Downwards
Frequency (Hz)
Time (ms)

Cortical Decomposition

- Spectrogram is frequency versus time
- Filter with various scale-rate combinations
  - Complex filter (details in the paper)
- Obtain a four-dimensional representation
  - Frequency, time, scale, and rate (and phase)
  - Called "cortical decomposition"
  - Simulates the sound representation used by the brain

Sample Scale-Rate Plots

- (here summed over all frequencies)



Usefulness

- Linear decomposition
  - Invertible
  - Predictable
- Used recently in:
  - Prediction of neural response
  - Evaluation of speech intelligibility
  - Separation of pitch and timbre

Inversion

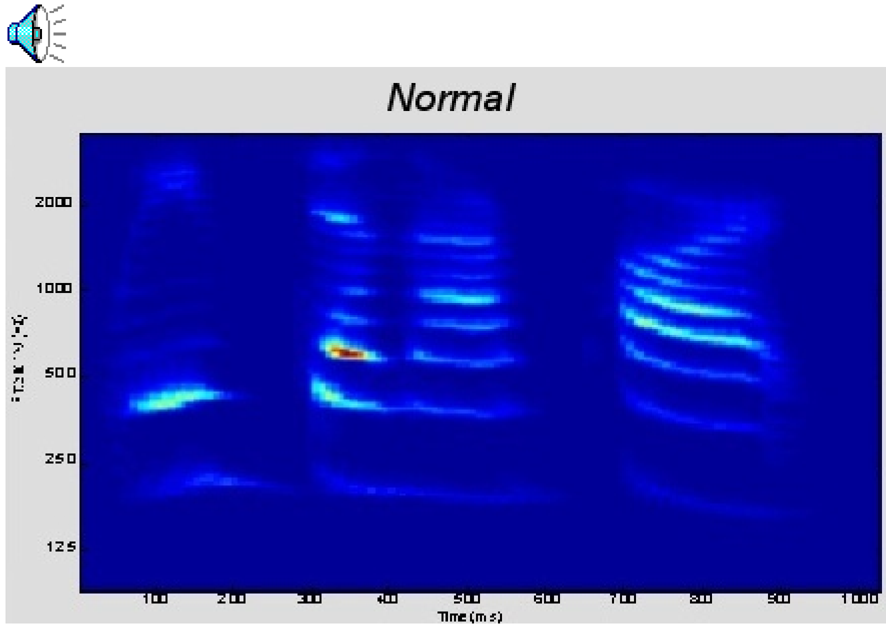
- Modify sound in cortical representation
- Compose the auditory spectrogram back
  - Linear, easy process
  - Think of it as an inverse Fourier transform
- Compose the signal back
  - Non-linear, hard process
  - Iterative solution (see paper)

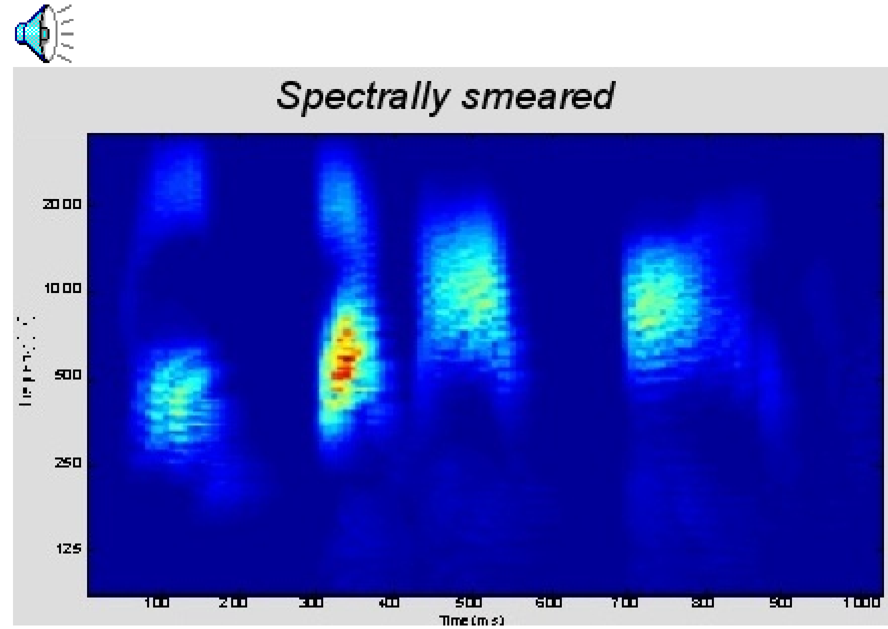
Fourier Transform Analogy

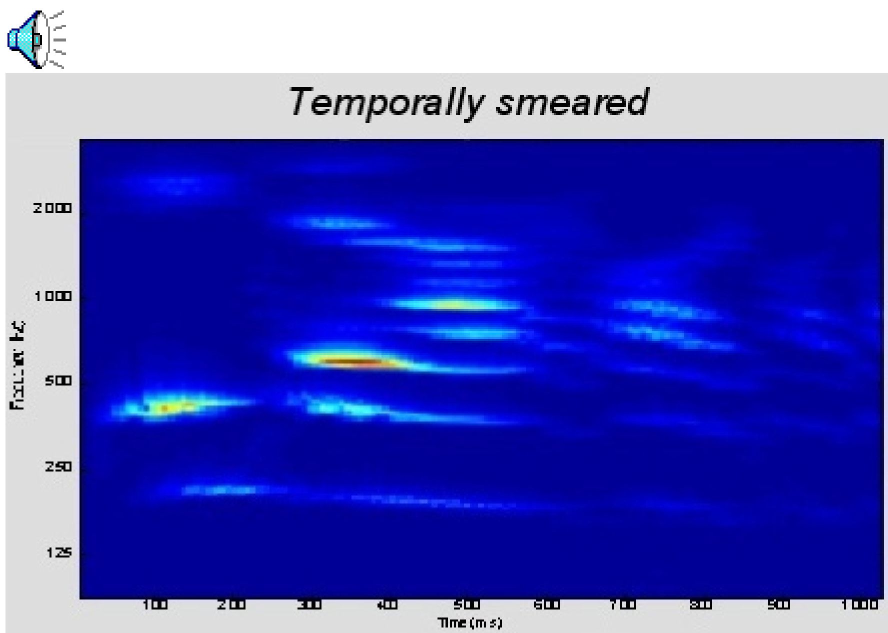
- Basis functions F(t): sin(Nωt) and cos(Nωt)
- Fourier transform:
  - Coefficient for F(t) shows the correlation (a measure of similarity) between the signal and F(t)
- Inverse Fourier transform:
  - Assemble the signal as a sum of all F(t) weighted by their appropriate coefficients

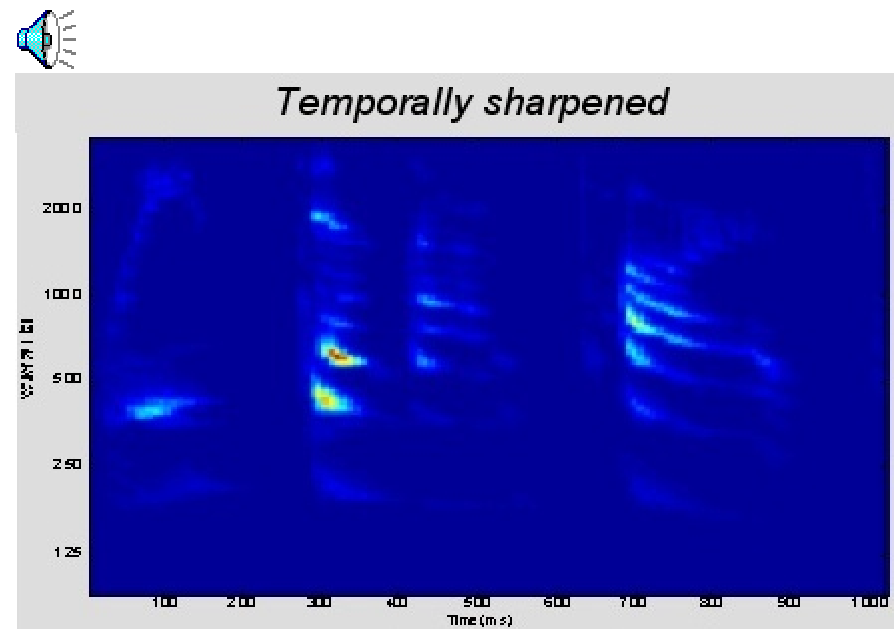
Fourier Transform Analogy

- Same with cortical representation, but…
  - 2-D input signal (instead of 1-D in FT)
  - Basis functions are of 2 parameters (rate and scale) (instead of one N in FT)
- Now can make some changes in cortical representation


Normal
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (ms)

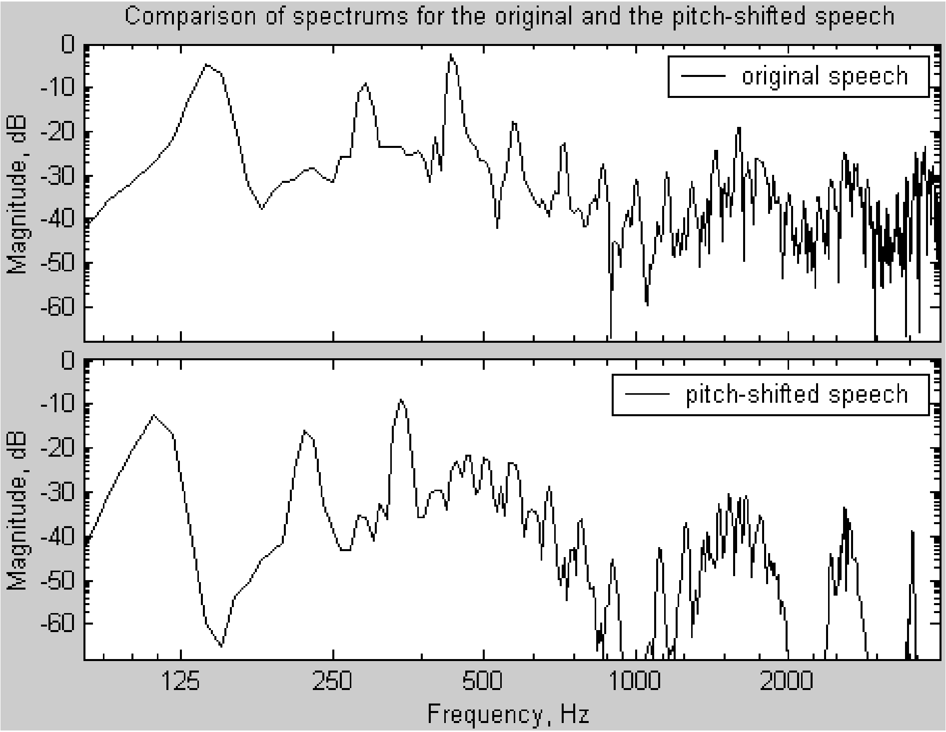

Spectrally smeared
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (ms)


Temporally smeared
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (ms)


Temporally sharpened
2000
1000
500
250
125
100 200 300 400 500 600 700 800 900 1000
Time (ms)

Application Example

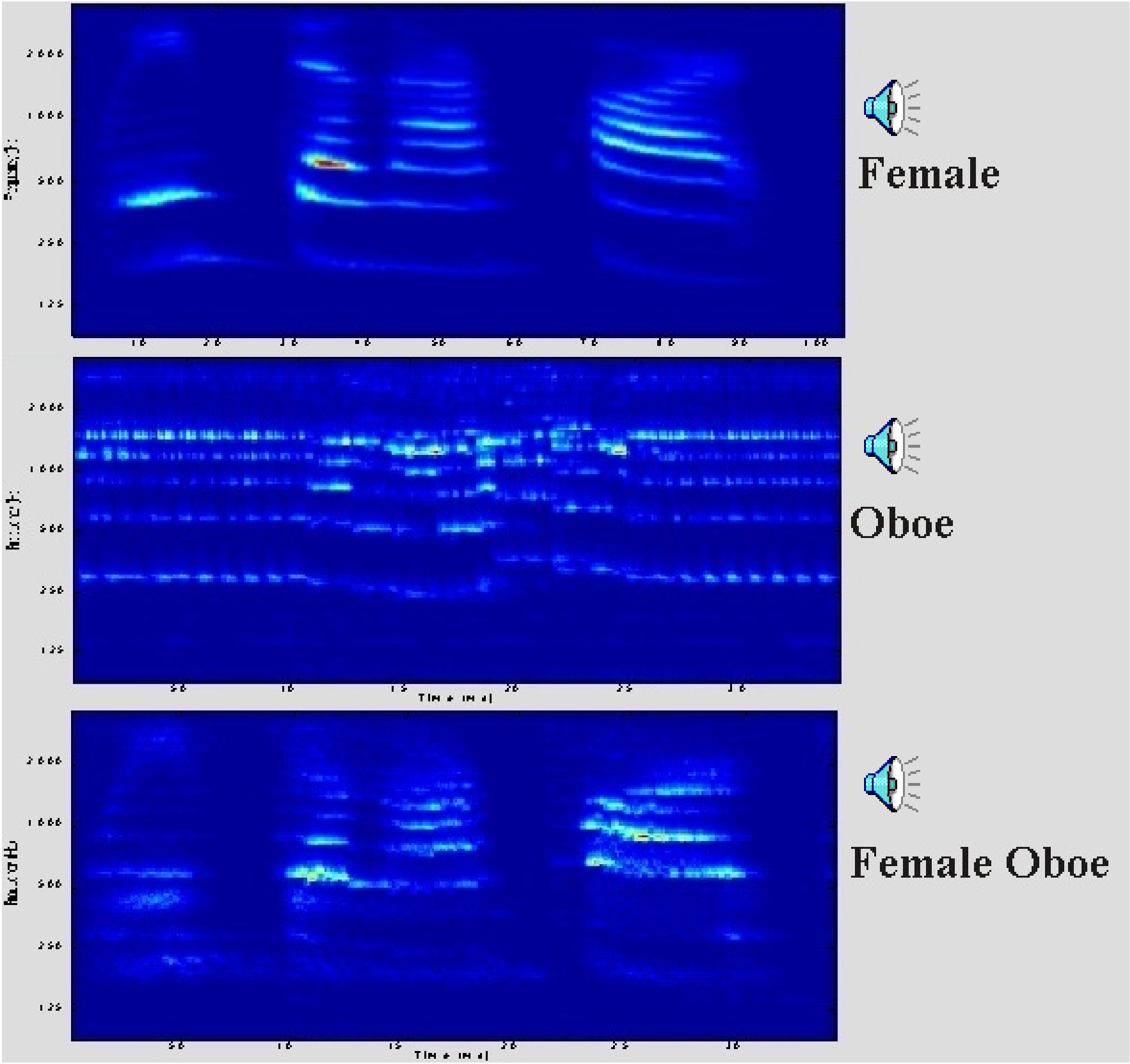
- Separation of pitch and timbre
- Selective modifications of either
- Narrow spectral features constitute pitch
- Wide spectral features (spectral envelope) is timbre



Application Example

- Separable in cortical representation
  - Can change one without changing the other
  - Can interpolate timbre
  - Can combine pitch of one person and timbre of another one
  - Or pitch of a person and a timbre of a musical instrument

Speaking Oboe



References

- http://www.isr.umd.edu/CAAR/ (code)
- http://pirl.umd.edu/NPDM/ (sound samples)
- "Neuromimetic sound representation for percept detection and manipulation", D. N. Zotkin, T. Chi, S. A. Shamma, and R. Duraiswami, EURASIP Journal on Applied Signal Processing, vol. 2005(9), pp. 1350-1364 (full description and more references).