LING 490 - SPECIAL TOPICS IN LINGUISTICS

Fundamentals of Digital Signal Processing

Yan Tang

Department of Linguistics, UIUC

Week 16

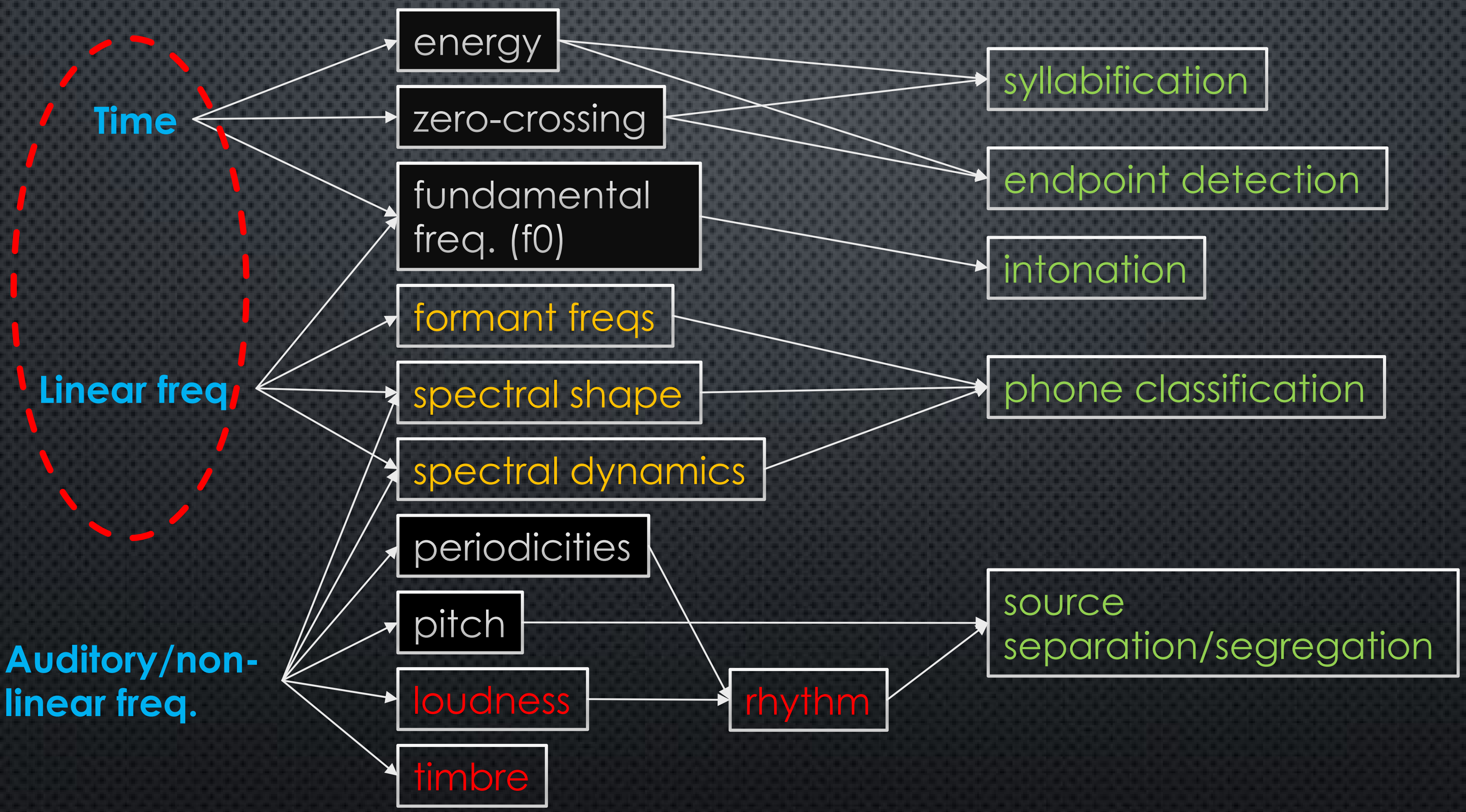
Last week…

- Spectral centre of gravity (CoG) / spectral centroid
  - A measure of the distribution of energy at different frequencies in a spectrum
  - The requisite for computing spectral moments and other measures
  - Related to many aspects of speech perception

Last week…

- Linear predictive coding (LPC)
  - Motivation: to separate the sound source and the filter effect, i.e. the vocal tract effect
  - The filter: from previous samples to estimate LPC coefficients; remove the effect from the source
  - The source: inverse filtering; remove the effect from the vocal tract
  - Applications: e.g. formant and F0 estimation, speech encoding, speech synthesis…

Useful parameters of speech signals



Cepstral processing: additive noise - motivation

- Signal analysis commonly concerns the separation of combined signal
  - Example: noise reduction, source separation
- Motivation: additive noise
  - Speech + noise: $x(t) = s(t) + n(t)$
- If we know something about the noise (bandlimited?), could transform to frequency domain, do some surgery, then return to time domain
- Surgery attempts to remove $n(t)$

Cepstral processing: convolution

source → filter → signal

- Recall source-filter model. How do we untangle this (by not using LPC!)

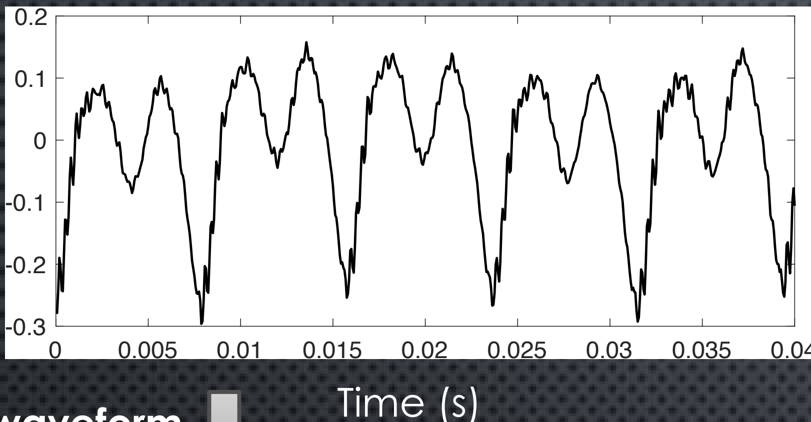
- Problem: unlike previous slide, source and filter are not added, but convolved:

$$\underbrace{x(t)}_{\text{speech signal}} = \underbrace{e(t)}_{\text{source}} \otimes \underbrace{h(t)}_{\text{filter}}$$

Cepstral processing: convolution

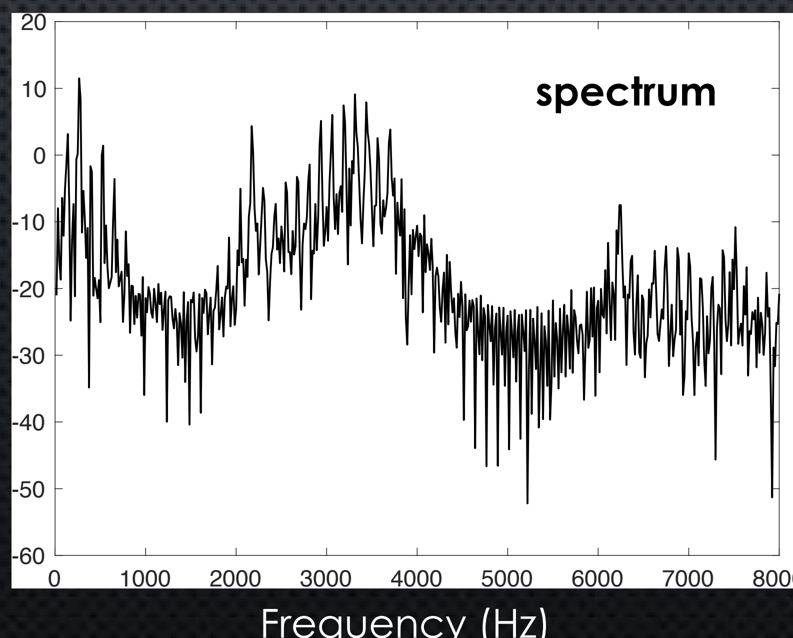
- One strategy is to convert convolution to addition and then use same approach as the speech+noise problem.
- How?
  - 1. Fourier transform converts convolution to multiplication
  - 2. Taking a log converts multiplication to addition
- These two steps are the basis of cepstral analysis

Pictorial justification



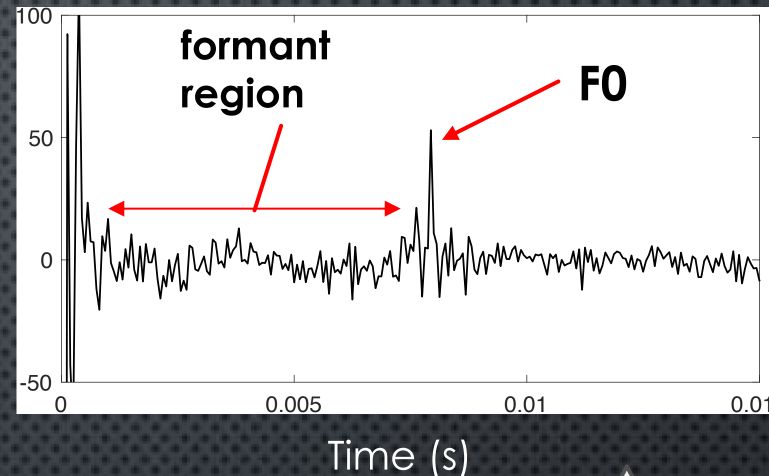
waveform

DFT
Log magnitude



inverse DFT

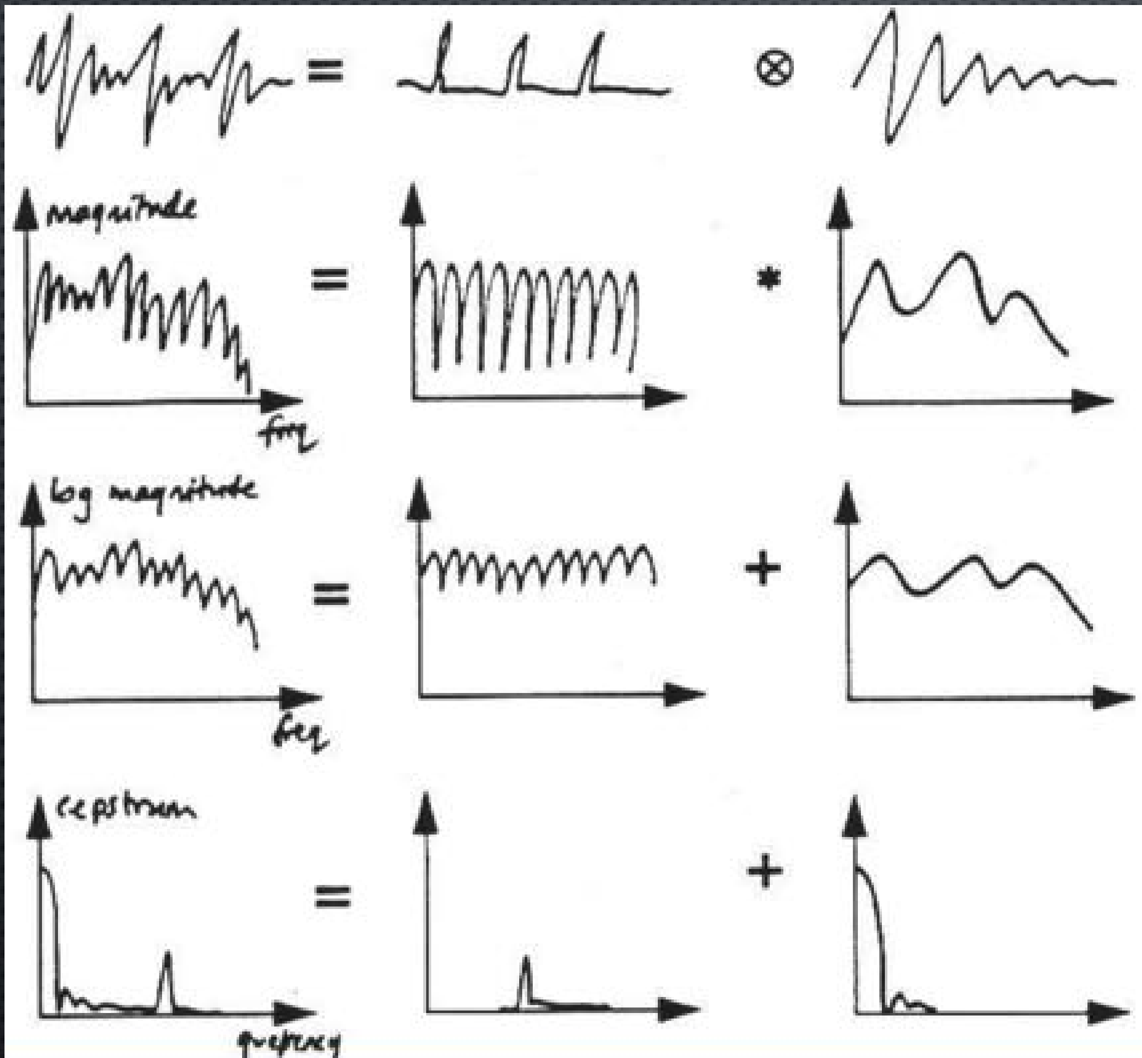
cepstrum



Definition

- The ***cepstrum*** is the inverse DFT of the log magnitude spectrum of a signal.
- We use the inverse DFT because we are transforming *from* the frequency domain

Method



- Fourier Transform (FT) turns convolution into multiplication
- Logs turn multiplication into addition
- FT is a linear transform, so added components remain added

Let's see that in equations

$$x(t) = e(t) \otimes h(t)$$

DFT

$$X(f) = E(f) \times H(f)$$

magnitude

$$|X(f)| = |E(f) \times H(f)|$$

log

$$\log(|X(f)|) = \log(|E(f) \times H(f)|) = \log\big(E(f)\big) + \log(H(f))$$

inverse DFT

$$F^{-1}[\log(|X(f)|)] = F^{-1}[\log\big(E(f)\big)] + F^{-1}[\log\big(H(f)\big)]$$

Terminology

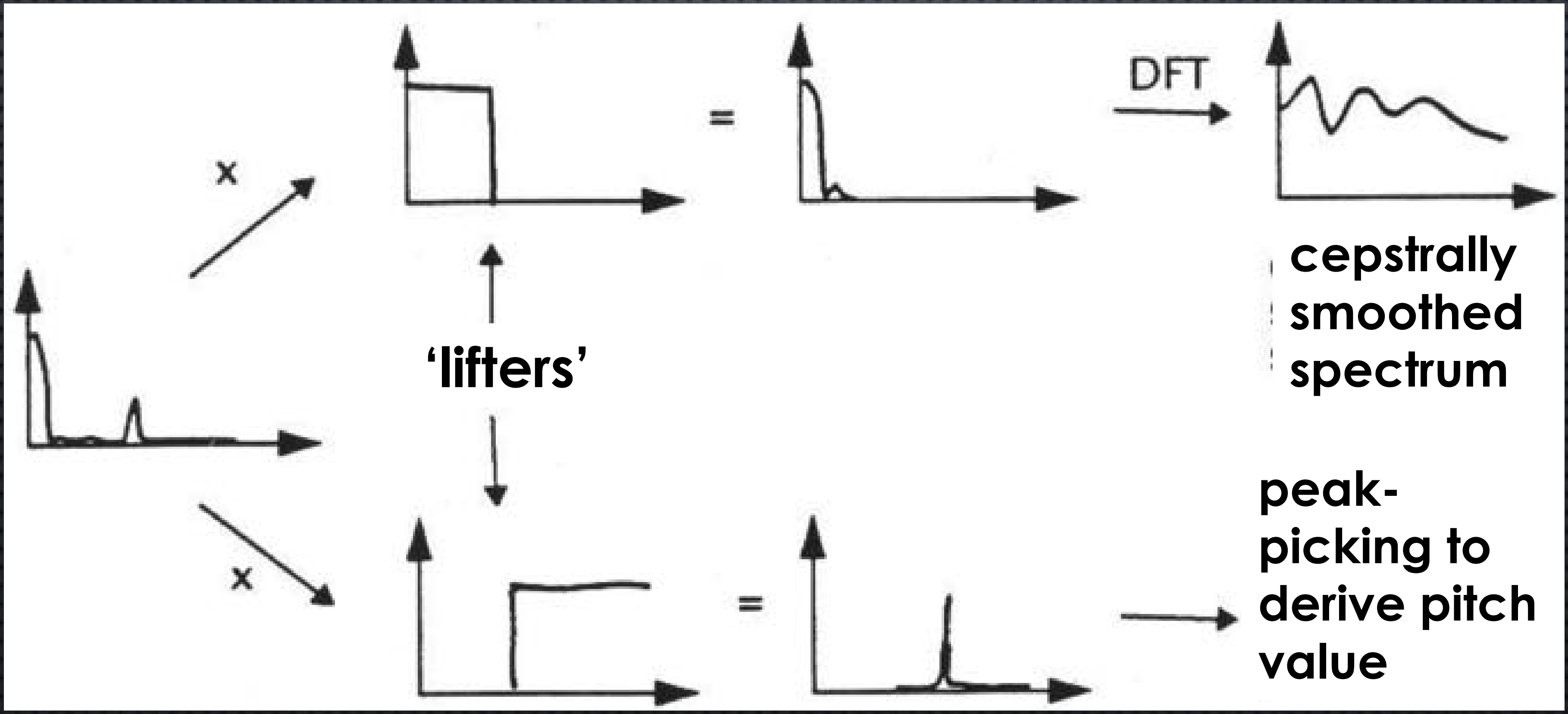
- What domain is the cepstrum in?
- Not quite the time-domain because we took a log magnitude between the DFT and inverse DFT
- A new domain!
- New terminology…

Terminology

- Spectrum → Cepstrum
- Frequency → Quefrency
- Filtering → Liftering
- etc

- Therefore, the domain itself is called the ***quefrency*** domain

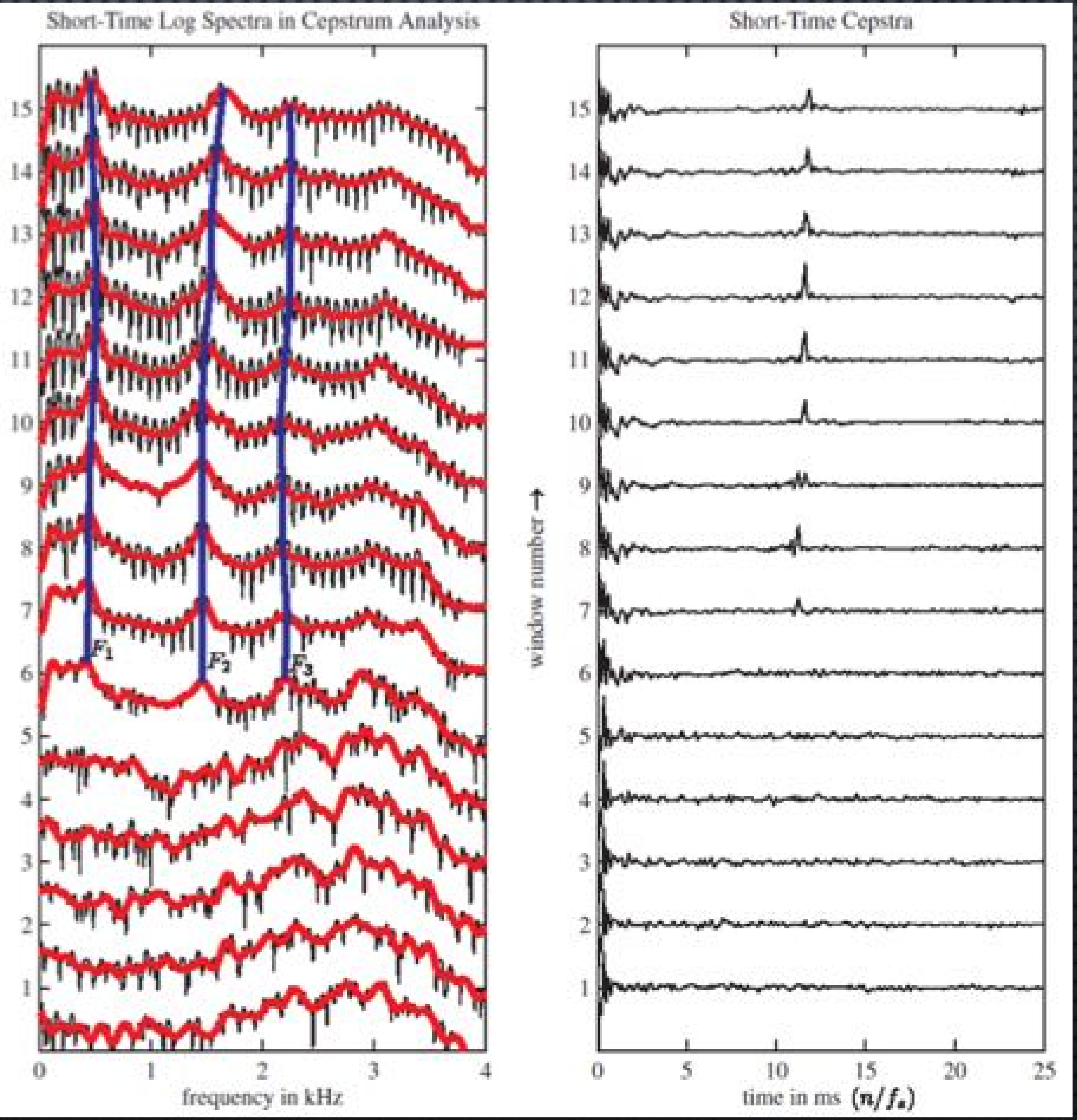
Untangling by liftering



Practical uses

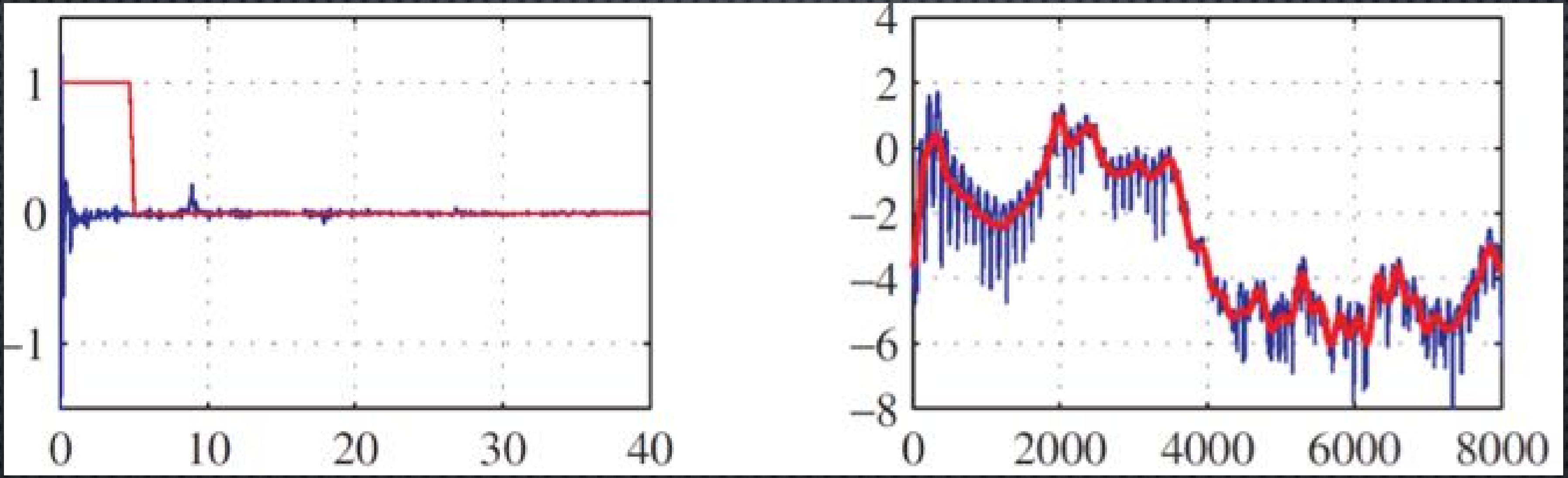
- ‘Pitch’ analysis
  - Locate and track the prominent cepstral peak after liftering
- Spectral smoothing
  - Remove the high quefrency region (containing the pitch peak) and invert.
- Low order cepstral coefficients
  - Very efficient (largely non-redundant) representation of the spectrum

Pitch analysis and formant tracking



Rabiner & Schafer, 2007

Spectral smoothing by liftering



Rabiner & Schafer, 2007

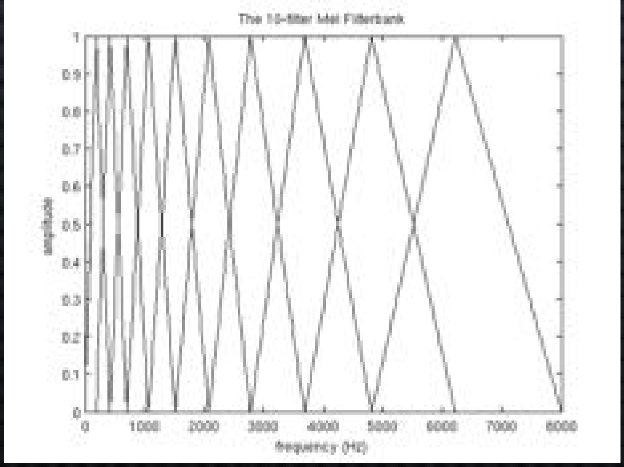
Mel Frequency Cepstral Coefficients (MFCCs)

- Probably most widely used features for speech recognition
- Actually just the first few cepstral coefficients (13 is not uncommon)
- Captures most of the **formant** information
- Produced in similar way as we've described except the frequency axis is warped prior to the final inverse DFT (not exactly!) to more closely match human pitch perception (the Mel scale)

Computing MFCCs

MFCC calculation is a frame-based process:

- Perform the short-term Fourier transform (STFT)
- Calculate the periodogram estimate of the power spectrum
- Apply the Mel filterbank to the power spectrum. Calculate the total energy in each band by summation
- Take the log of the energy of each band
- Perform discrete cosine transform (DCT) on the log energies



Remarks on MFCC and cepstrum analysis

- The MFCCs are calculated on the Mel scale – a perceptual scale for pitch judgement
  - Match human pitch perception
- The MFCC is no longer a homomorphic transformation
  - Due to the log operation
- MFCC uses the DCT instead of the IDFT
  - Decorrelate the Mel-scale log energies
  - MFCCs are good features for machine learning