Feature extraction 1

Dr Philip Jackson

- Cepstral analysis
  - Real & complex cepstra
  - Homomorphic decomposition
- Filter bank energies
- Mel-frequency cepstral coefficients

Cepstral analysis (1)

Cepstral analysis, or "homomorphic decomposition", is designed to separate convolved signal components by transforming the signal s to a domain where components are additive and in distinct regions.

A source signal x passed through a filter with impulse reponse h, where $\otimes$ denotes convolution, yields

$$s(t) = h(t) \otimes x(t) \tag{1}$$

Taking Fourier transforms of both sides, we get

$$S(\omega) = H(\omega)\, X(\omega) \tag{2}$$

where ω denotes the angular frequency, $\omega = 2\pi f$.

Cepstral analysis (2)

So, the spectrum of the signal can be written

$$S(\omega) = H(\omega)\ X(\omega)$$

and taking natural logarithms of both sides gives

$$\ln S(\omega) = \ln H(\omega) + \ln X(\omega) \tag{3}$$

Thus, a convolution in time has been transformed into a sum of log components in the frequency domain.

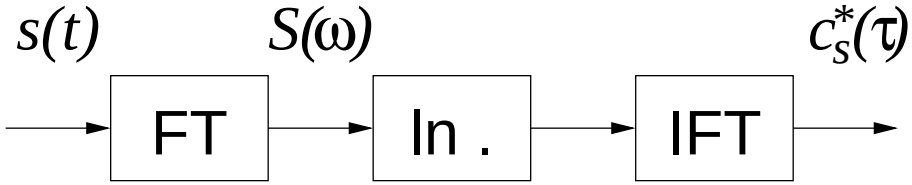
Finally, applying an inverse Fourier transform to the log spectrum gives

$$\mathcal{F}^{-1}\{\ln S(\omega)\} = \mathcal{F}^{-1}\{\ln H(\omega)\} + \mathcal{F}^{-1}\{\ln X(\omega)\} \tag{4}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform (FT), and $\mathcal{F}^{-1}$ its inverse (IFT).

Cepstral analysis (3)

A block diagram of the process for calculating cepstral coefficients shows the three steps:



This last transform takes the function back into the time domain, but it is *not* the same as the time of the original signal. In fact, it is a measure of the rate of change of the spectral envelope, as viewed in decibels (dB).

This domain is called the **cepstrum**, and the time axis is often referred to as the *lag* or "*quefrency*" axis.

Complex cepstrum

In this case, the output $c_s^*(\tau)$ is the **complex cepstrum** of signal s, since the logarithm is applied to a complex number. Hence, we can re-write eq. 4 as

$$c_s^*(\tau) = c_h^*(\tau) + c_x^*(\tau) \tag{5}$$

where, for example,

$$c_s^*(\tau) = \mathcal{F}^{-1}\{\ln S(\omega)\} \tag{6}$$

and $c_h^*(\tau)$ and $c_x^*(\tau)$ are the complex cepstra of h and x, respectively.

Real cepstrum

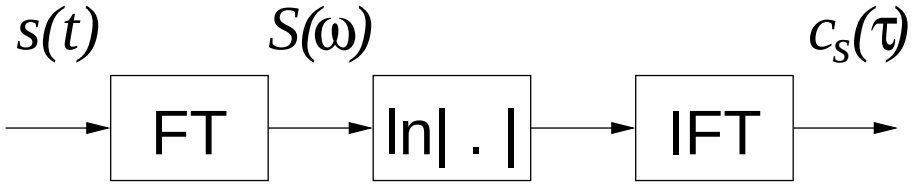
The **real cepstrum** is derived from the log magnitude of the spectrum,

$$c_s(\tau) = \mathcal{F}^{-1}\left\{\ln|S(\omega)|\right\} \tag{7}$$

and these are also are superposed:

$$c_s(\tau) = c_h(\tau) + c_x(\tau) \tag{8}$$

The diagram shows magnitude and logarithm operations:



The real cepstrum is a real and an even function of the lag, or quefrency. So, the IFT can be replaced by the discrete cosine transform (DCT).

Phase information from the original signal has been lost in the process of calculating the real cepstrum.

Real cepstrum applied to speech

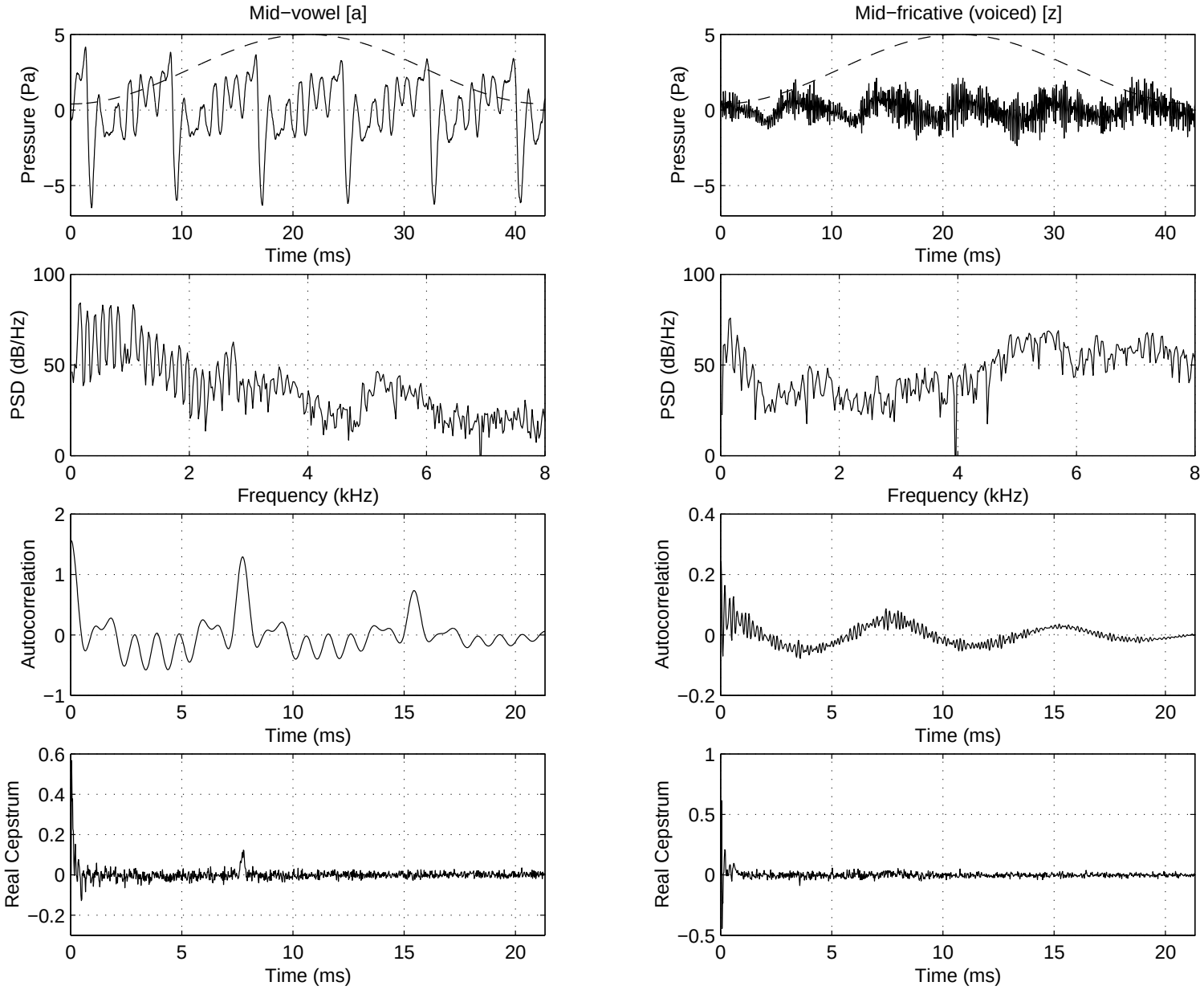
Cepstral analysis has found many applications in such areas as seismic exploration and speech processing, for which an example is given.

The sequence of plots below shows the cepstral analysis procedure applied to two frames of voiced speech data:

- a vowel [a] (vowels are high in amplitude and have strong periodicity),

- a voiced fricative consonant [z] (fricatives tend to have a strong high-frequency noise component).

Examples of cepstral analysis, [a] and [z]



Homomorphic decomposition

Decomposition by cepstral liftering

Provided that the *spectral properties* of the two signals, $h(t)$ and $x(t)$, are distinct, then $c_h(\tau)$ and $c_x(\tau)$ will occupy distinct regions of the quefrency domain.

Using a suitable cepstral filter (or *lifter*), the components may be separated from each other, and then they can be transformed back into log-magnitudes or magnitudes in the frequency domain, as required.

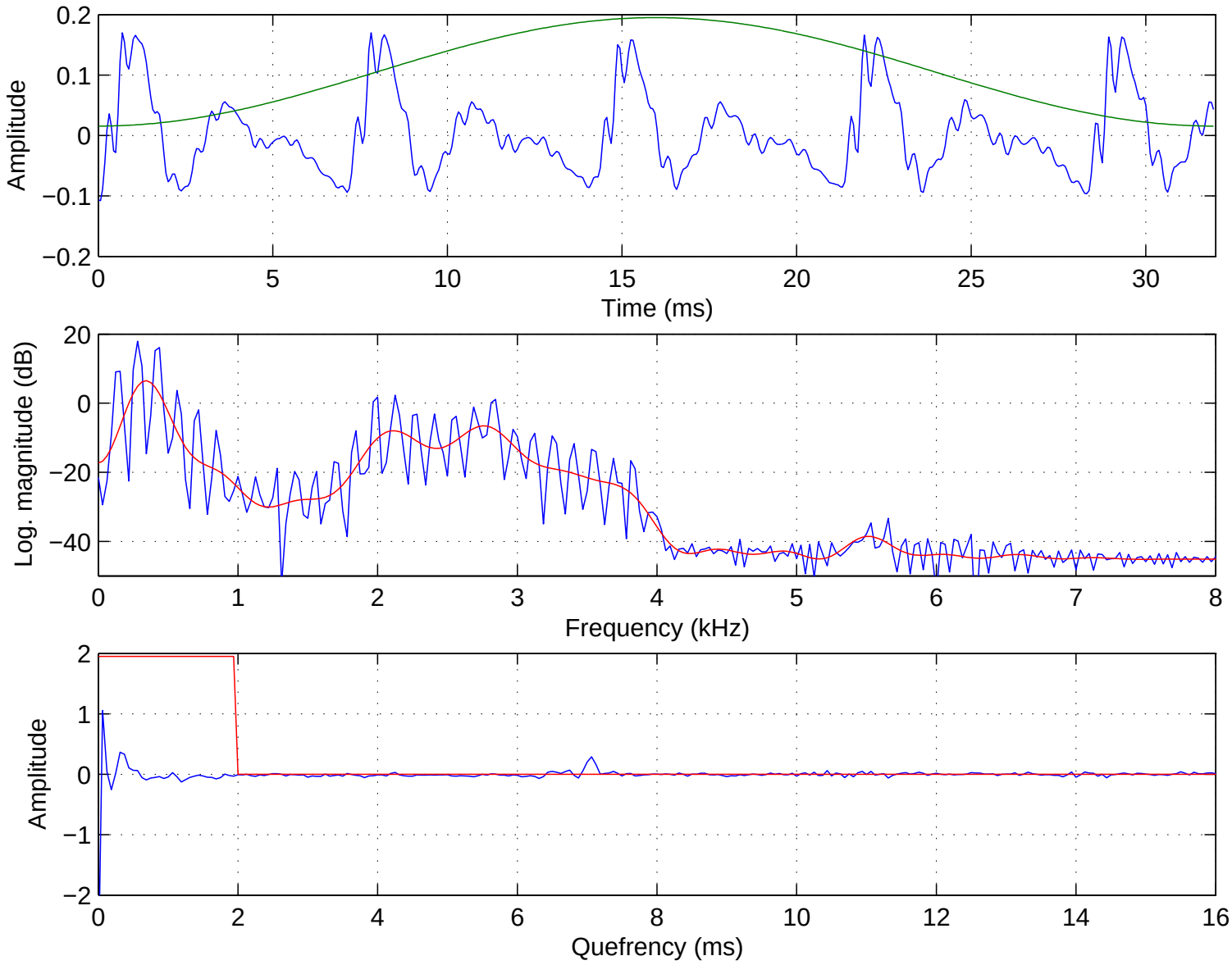
Source-filter theory of speech

It is assumed that the recorded speech signal is the output from a linear system which consists of a source of filter excitation (a series of periodic pulses) convolved with the impulse response of a filter (Fant 1960).

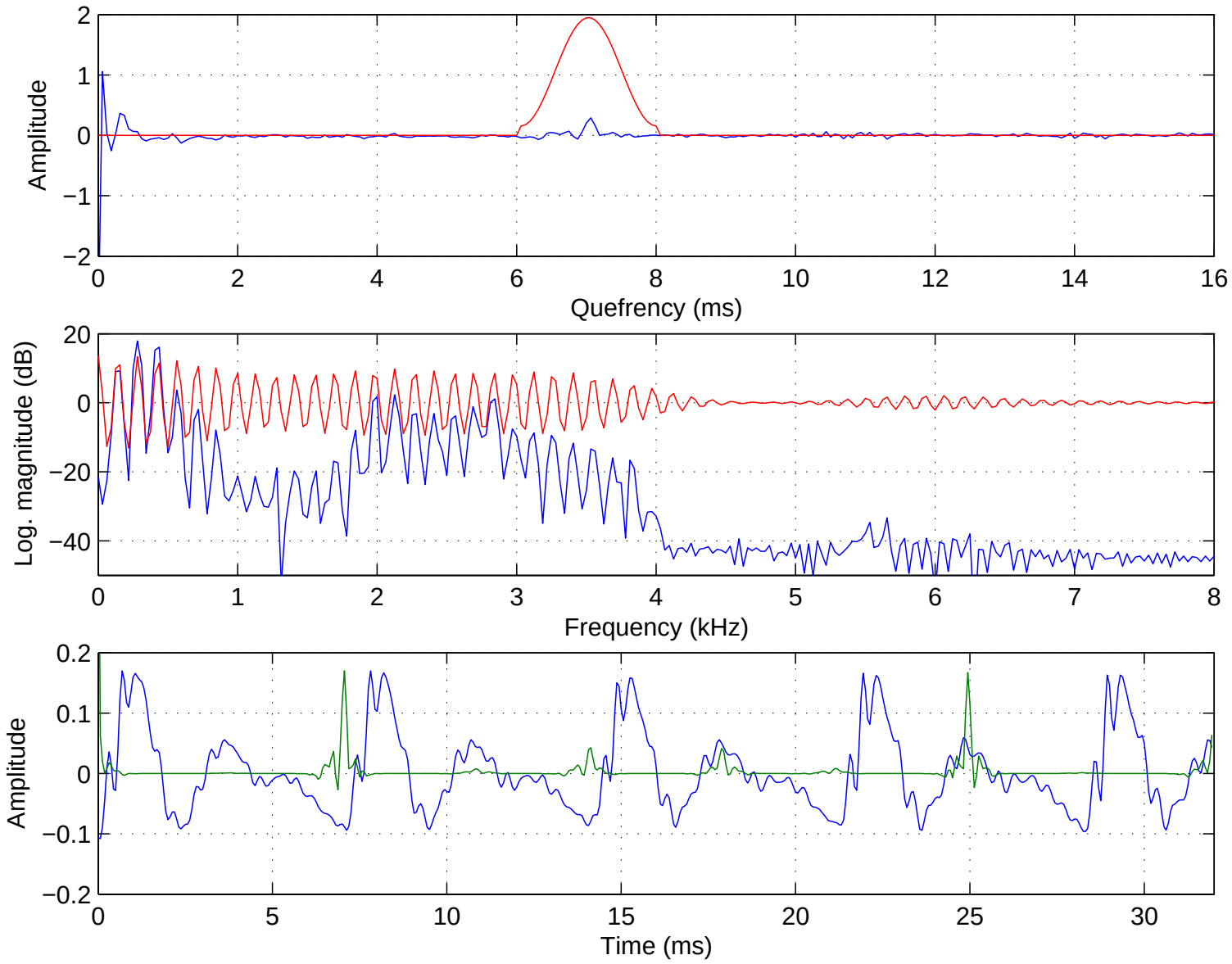
The filter represents the acoustic effect of the vocal tract, which depends on the positions of the articulators (jaw, tongue, lips, etc.) and corresponds to the uttered vowel ([i] from the word "linear").

Cepstral analysis allows both an accurate estimation of the periodicity of the excitation and the extraction of the frequency response, and hence the impulse response of the vocal-tract filter.

Example spectral envelope of [i] in “linear”



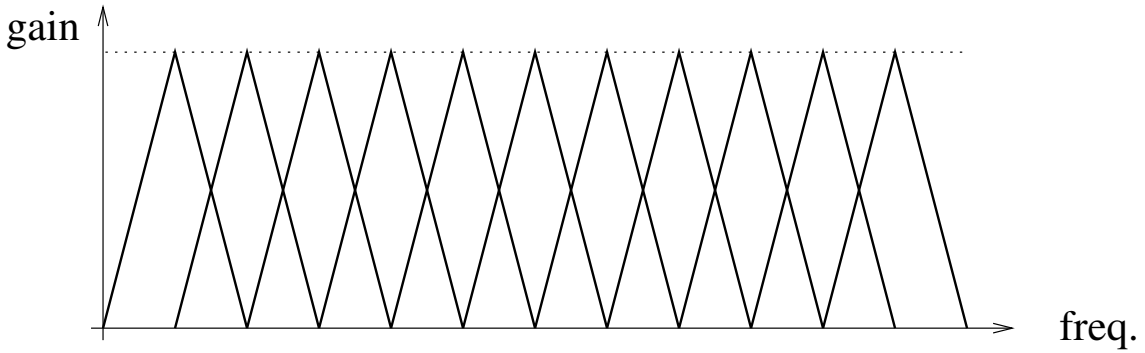
Example pitch extraction from [i] in “linear”



Extraction of acoustic features

Filter banks

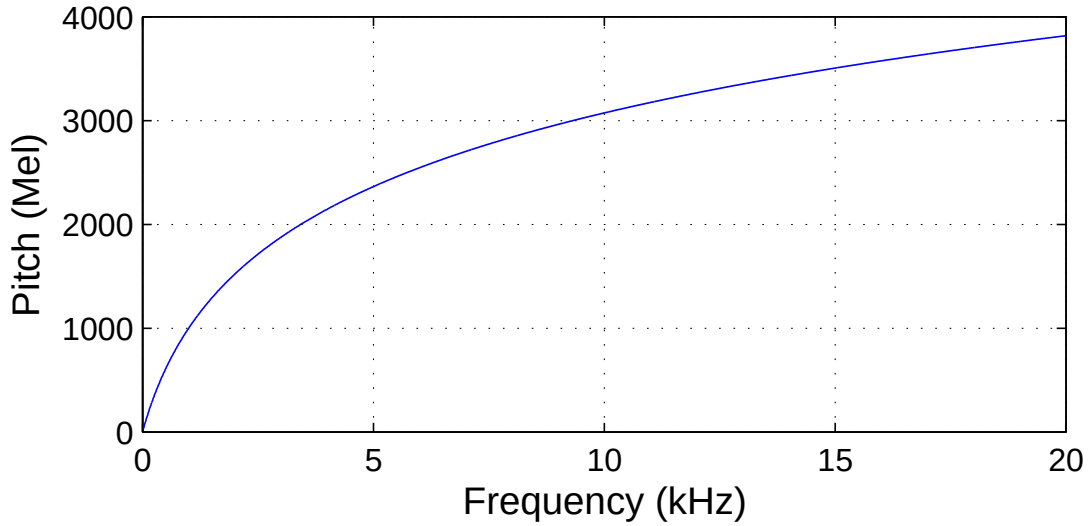
- measure local spectral shape
- reduce effects of pitch
- characteristic frequencies of nerve fibres
- filters with critical-band reponses



Mel-frequency scale

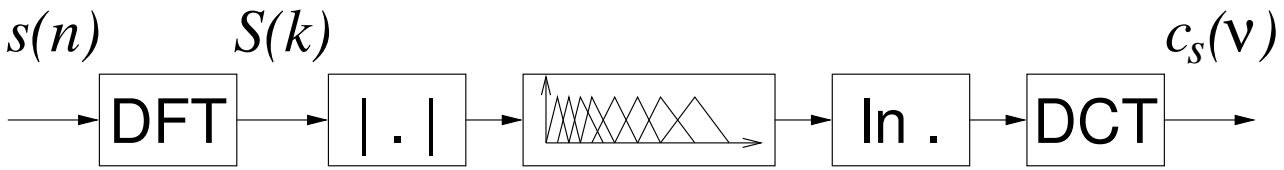
Human hearing does not perceive frequencies over $\sim$1 kHz in a linear fashion. The frequency resolution of the ear, measured in terms of the **critical bandwidth**, gets broader as frequency increases. Psychoacoustic experiments of this phenomenon have provided us with a normalised frequency scale, measured in Mels:

$$f_{\text{Mel}} = 1127.01048 \ln\left(1 + \frac{f_{\text{Hz}}}{700}\right) \tag{9}$$



Mel-frequency cepstral coefficients (MFCCs)

MFCC features are typically computed in the "front end" of an automatic speech recognition system:



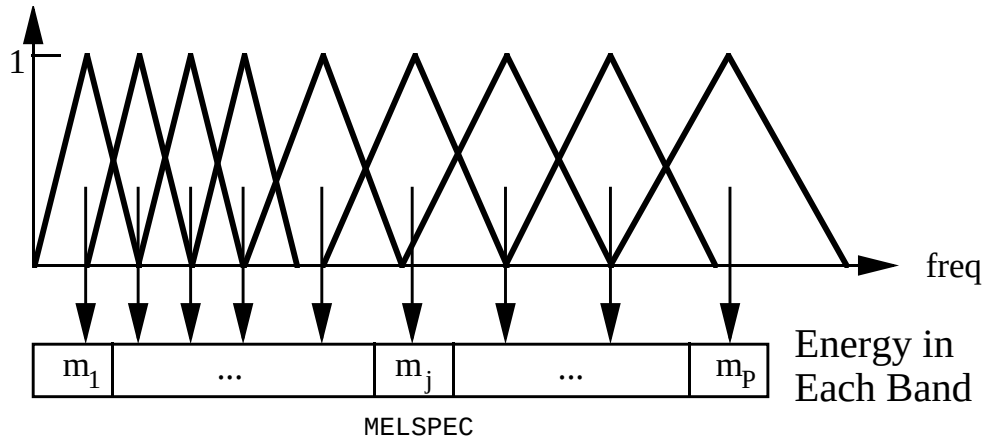


Figure 5.3. Mel-scale filter bank, from (Young et al, 1997).

Feature extraction 1 summary

- Cepstral analysis
  - Calculating the complex cepstrum
  - Calculating the real cepstrum
- Real cepstrum
  - Examples of cepstra computed from speech
- Homomorphic decomposition of speech
  - Spectral envelope
  - Pitch tracking
- Mel-frequency cepstral coefficients
  - Mel-frequency warping
  - As a front end for automatic speech recognition

Homework from week 3

- Complete the assignment for Lab 1
- Work through the exercises on the web page