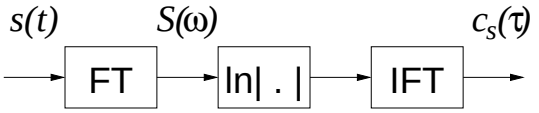
UniS Centre for Vision Speech & Signal Processing
University of Surrey, Guildford GU2 7XH.

Speech Processing

by Dr Philip Jackson

- Cepstral analysis
  - Real & complex cepstra
- Homomorphic decomposition



http://www.ee.surrey.ac.uk/Personal/P.Jackson/ce.adsp/

Cepstral analysis (1)

Sometimes referred to as "homomorphic decomposition", this technique is designed to separate convolved signal components by transforming the signal into a domain where the convolution become a simple summation. Let

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

where $\otimes$ denotes convolution. Then, taking Fourier transforms of both sides, we have

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

where the uppercase variables represent the complex spectra of the lowercase variables in time.

Cepstral analysis (2)

The magnitude (or root-power) spectrum of the signal can be written

$$|S(\omega)| = |X(\omega)|\,|Y(\omega)|\,, \tag{3}$$

and taking logarithms of both sides gives

$$\ln|S(\omega)| = \ln|X(\omega)| + \ln|Y(\omega)|\,. \tag{4}$$

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

One final stage is required if we want to separate the x and y components.

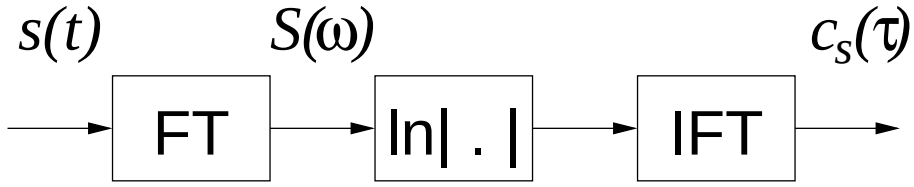
However, it should be noted that the phase information from the original signal has been lost, as a result of the magnitude operation on the complex spectrum.

Cepstral analysis (3)

Applying the inverse Fourier transform to the log spectrum gives

$$\mathcal{F}^{-1}\{\ln|S(\omega)|\} = \mathcal{F}^{-1}\{\ln|X(\omega)|\} + \mathcal{F}^{-1}\{\ln|Y(\omega)|\}, \quad (5)$$

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



This last transform takes the signal back into a time domain representation, but *not* the same as the time axis of the original waveform; in fact, it is a measure of the rate of change of the spectral magnitudes. This domain is called the *cepstrum*, and the time axis is often referred to as the *lag* or "*quefrency*".

Real cepstrum

Letting $c_s(\tau)$, $c_x(\tau)$ and $c_y(\tau)$ be the cepstra of signals s, x and y, respectively, it can be seen that

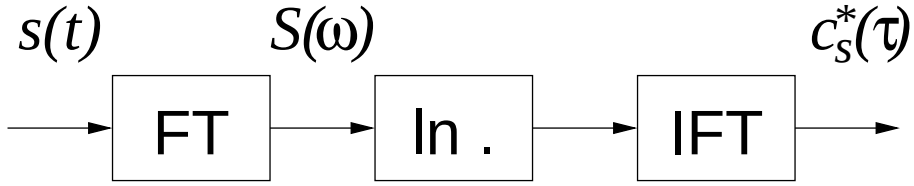
$$c_s(\tau) = c_x(\tau) + c_y(\tau). \tag{6}$$

This function is called the *real cepstrum* because it is derived from the power spectrum of the signal, which is always a real function of frequency. The cepstrum is also real and is an even function of the independent variable, lag or quefrency.

Note that, because the log-magnitude spectrum is real and symmetrical (i.e., even) for real signals, the final IFT can be replaced with a cosine transform.

Complex cepstrum

One difficulty with the real cepstrum concerns the loss of phase information. However, a similar quantity can be formed without the magnitude operation, and hence using the complex logarithm:



As before, the signals' *complex cepstra*, $c_s^*(\tau)$, $c_x^*(\tau)$ and $c_y^*(\tau)$, are superposed,

$$c_s^*(\tau) = c_x^*(\tau) + c_y^*(\tau). \tag{7}$$

Real cepstrum applied to speech

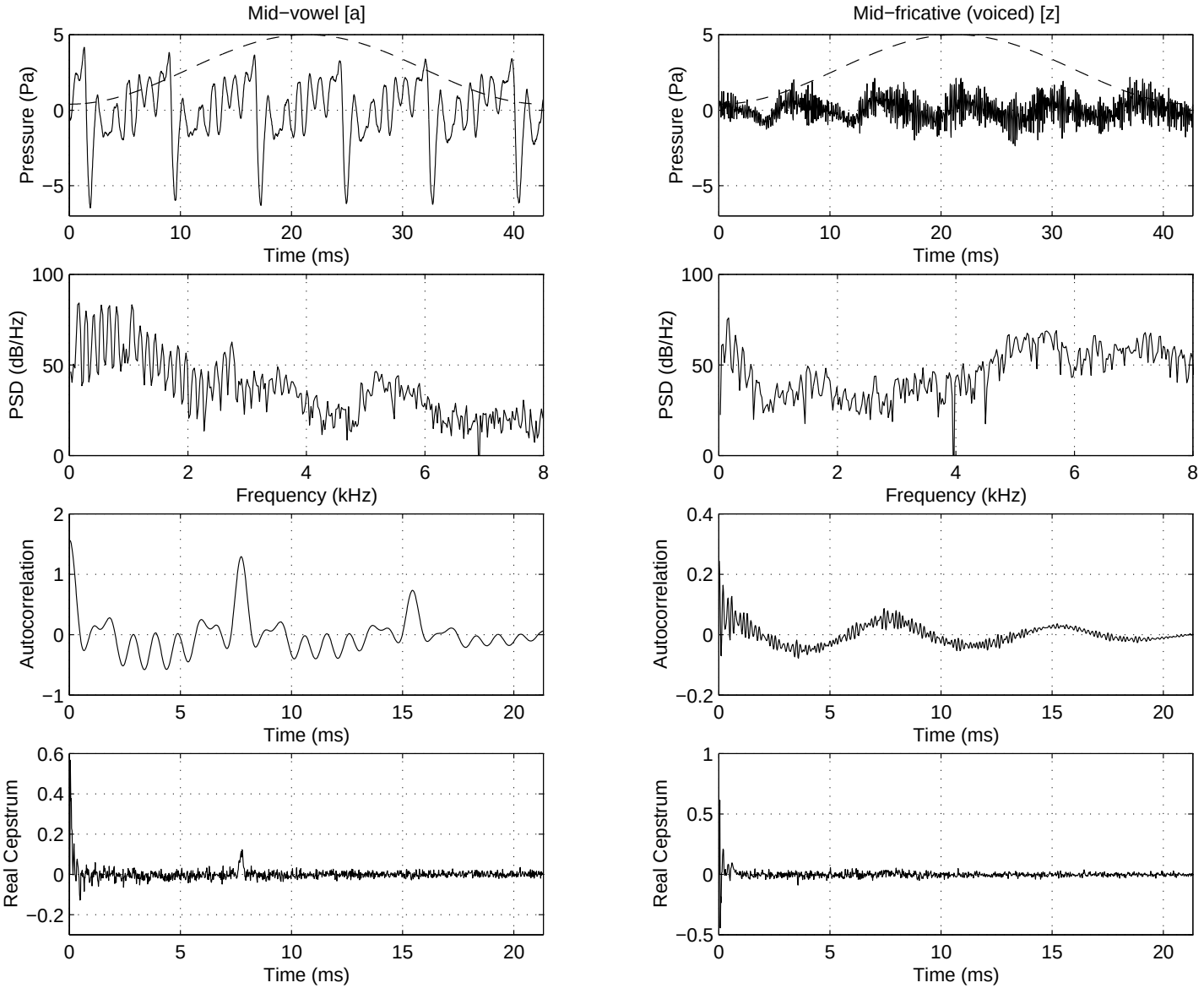
Cepstral analysis has found many applcations in such areas as seismic exploration and speech processing, for which an example is given.

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

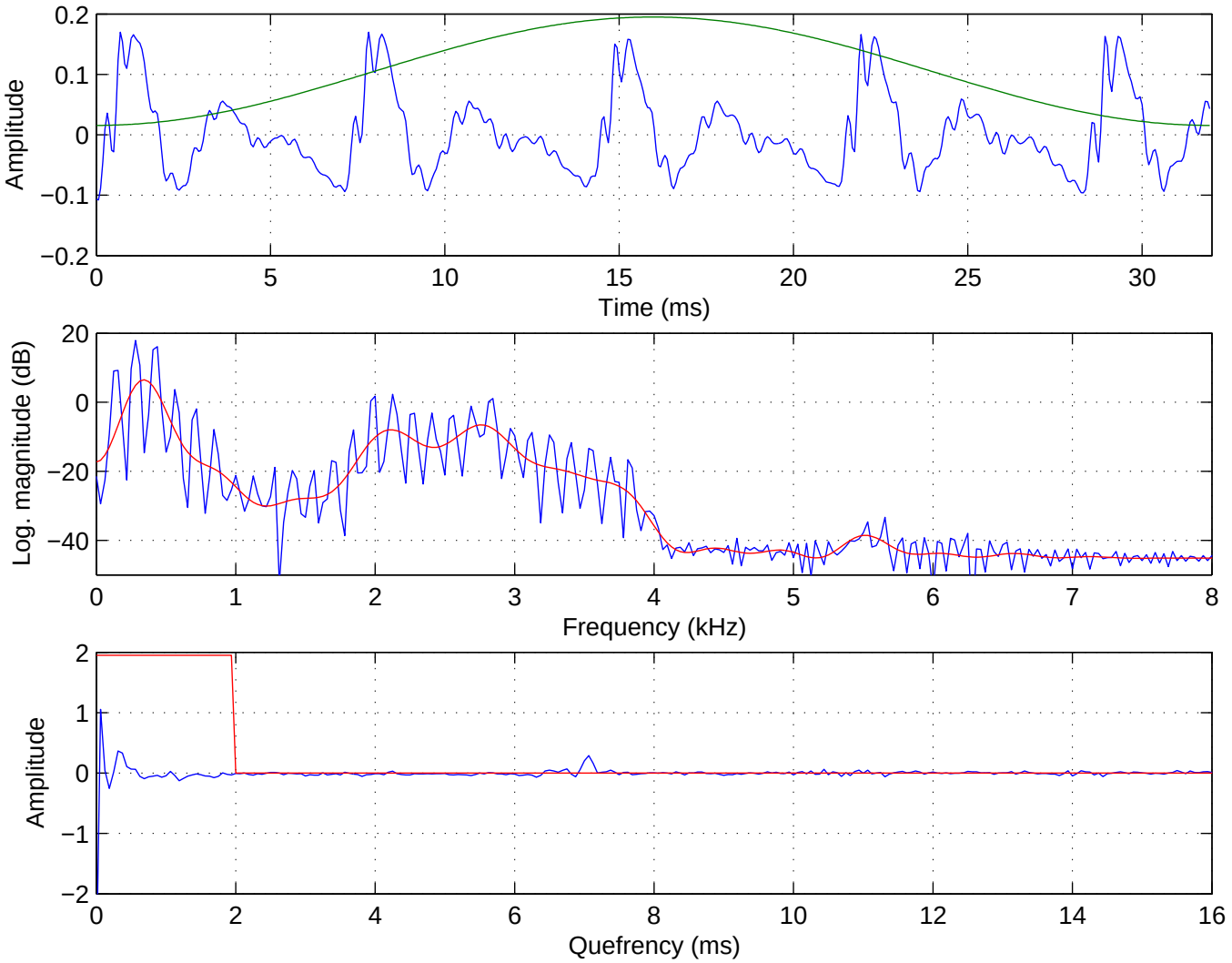
Provided that the *spectral properties* of the two signals, $x(t)$ and $y(t)$, are distinct, then $c_x(\tau)$ and $c_y(\tau)$ will occupy distinct regions of the quefrency domain. Using a suitable cepstral filter (or *lifter*), the components may be separated from each other, and then they can be transformed back into log-magnitudes or magnitudes in the frequency domain, as required.

Source-filter theory of speech

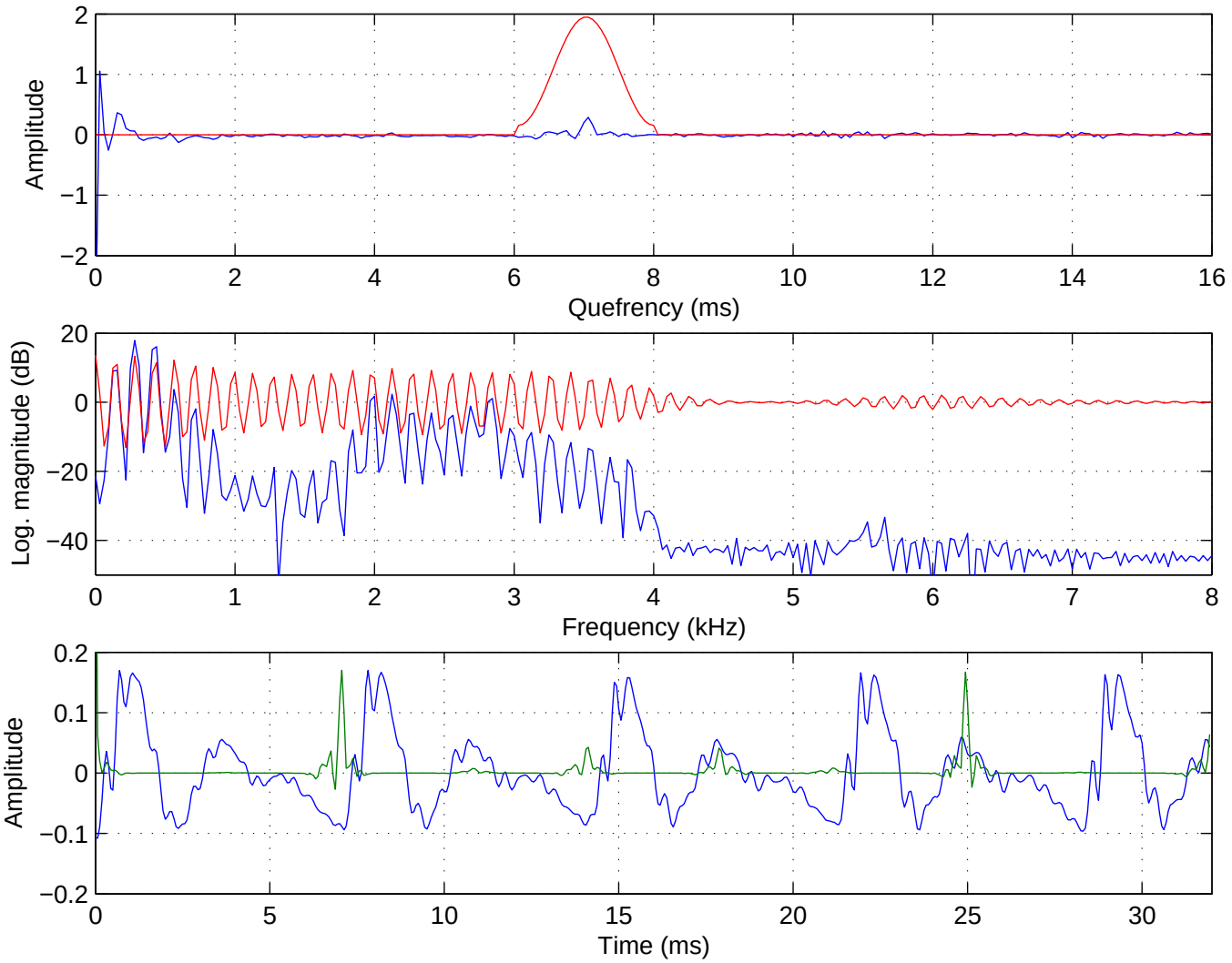
It is assumed that the recorded speech signal is the output from a linear system which consists of a source of filter excitation (a series of periodic pulses) convolved with the impulse response of a filter (Fant 1960). The filter represents the acoustic effect of the vocal tract, which depends on the 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”



Summary

- Cepstral analysis
  - Calculating the real cepstrum
  - Calculating the complex cepstrum
- Real cepstrum
  - Examples of cepstra computed from speech
- Homomorphic decomposition of speech
  - Spectral envelope
  - Pitch tracking