COM3502-4502-6502
SPEECH PROCESSING

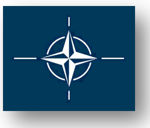
Lecture 18
Linear Prediction

1

Linear Prediction

'Linear prediction' is the basis for almost all modern speech coding algorithms …

- mobile phones
- internet telephony (*VOIP*)
- military communications



2

Linear Prediction Analysis

- Linear prediction assumes that each sample (*of a speech signal*) can be predicted from a weighted sum of the p preceding samples

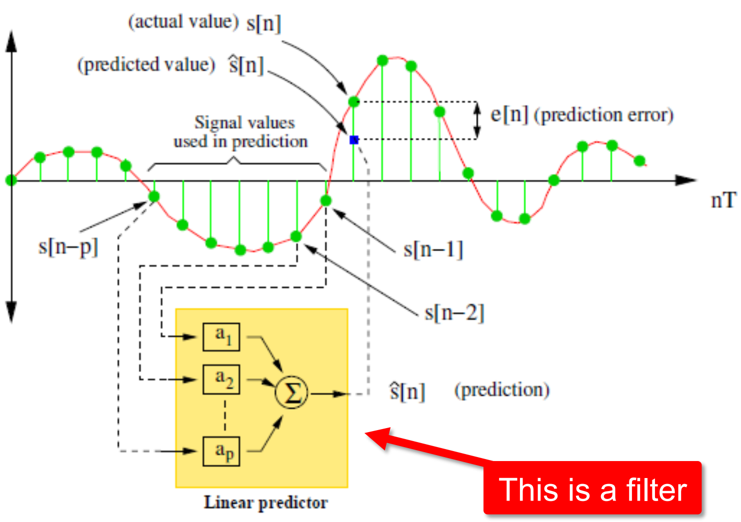
$$\hat{s}[n] = a_1 s[n-1] + a_2 s[n-2] + \ldots + a_p s[n-p]$$

$$= \sum_{i=1}^{p} a_i s[n-i]$$

- Exploiting redundancy arising from the *stationarity* assumption, linear prediction allows a speech frame of several hundred samples to be represented by only 10-15 'prediction coefficients' ($a_1 \ldots a_p$)

The University Of Sheffield.

Linear Prediction Analysis



This is a filter

The University Of Sheffield.

Linear Prediction Analysis

- The prediction error for a particular sample *e[n]* is …

$$e[n] = s[n] - \hat{s}[n] = s[n] - \sum_{i=1}^{p} a_i s[n-i]$$

- Taking Z-transforms of both sides we obtain …

$$E(z) = \left[1 - \sum_{i=1}^{p} a_i z^{-i}\right] S(z)$$

- Hence the filter transfer function is …

$$H(z) = \frac{E(z)}{S(z)} = 1 - \sum_{i=1}^{p} a_i z^{-i}$$

An 'all-zero' filter

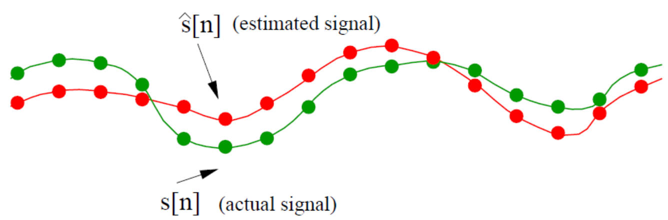
The University Of Sheffield.

5

Calculating Prediction Coefficients

- The linear prediction coefficients are computed to minimise the sum-squared 'prediction error' ('*residual signal*') E_p over a frame of speech data

$$E_p = \sum_n e[n]^2 = \sum_n \left(s[n] - \sum_{k=1}^{p} a_k s[n-k] \right)^2$$



- This has the effect of 'whitening' the error signal (*i.e. flattening its spectrum*)

The University Of Sheffield.

6

Calculating Prediction Coefficients

- E_p is minimised by finding the solution to …

$$\frac{\partial E_p}{\partial a_k} = 0$$

- This results in a set of p simultaneous equations called the 'normal equations' …

$$\sum_n s[n]s[n-k] = \sum_{i-1}^{p} a_i \sum_n s[n-i]s[n-k] \qquad k = 1\ldots p$$

- Different ranges to the summation gives rise to …
  - the 'autocorrelation method'
  - the 'covariance method'

The University Of Sheffield.

Covariance Method

- In the covariance method, a windowed frame of N samples is considered from the outset
- By defining …

$$\phi_{i,k} = \sum_{n=0}^{N-1} s[n-i]s[n-k]$$

- The normal equations can be re-written as …

$$\phi_{0,k} = \sum_{k=1}^{p} a_i \phi_{i,k}$$

- These are solved using a general purpose matrix inversion routine for symmetric matrices
- Note that the consequence of the summation is that this method considers an overlap that extends into the previous speech frame

The University Of Sheffield.

Autocorrelation Method

- In the autocorrelation method, the summation is taken over all samples …

$$\sum_{n=-\infty}^{\infty} s[n-i]s[n-k] = \sum_{n=-\infty}^{\infty} s[n]s[n+i-k]$$

- This is the autocorrelation sequence r_{i-k}
- Hence the normal equations can be re-written as …

$$r_k = \sum_{i=1}^{p} a_i r_{i-k} \qquad k = 1 \dots p$$

- Note that, since the signal is non-stationary, it first needs to be windowed
- The normal equations are solved using a matrix inversion method such as 'Durban's algorithm'

The University Of Sheffield.

Comparison of Methods

	Autocorrelation	Covariance
Filter order	10-15	10-15
Windowing	✓	×
Stability	✓	×
Window size	25 msec	5-10 msec
Computation	+	+++
Reflection coeffs.	✓	×

The University Of Sheffield.

Linear Prediction

- The filter transfer function from the signal to the prediction error sequence is …

$$H_1(z) = \frac{E(z)}{S(z)} = 1 - \sum_{i=1}^{p} a_i z^{-i}$$

An 'all-zero' filter

- Hence the filter transfer function from the prediction error sequence to the signal is …

$$H_2(z) = \frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}}$$

An 'all-pole' filter

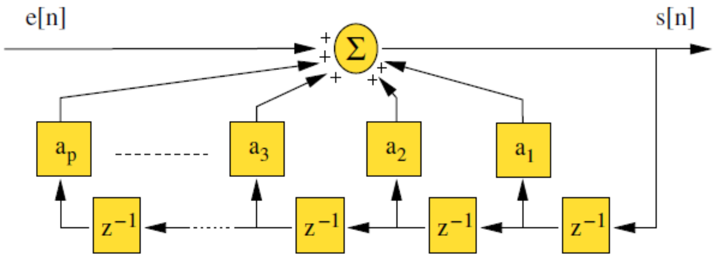
The University Of Sheffield.

11

© 2019 The University of Sheffield

Linear Prediction Synthesis

The LP filter can be configured as an all-pole digital filter with the error signal as its input …



$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}}$$

The University Of Sheffield.

12

Linear Prediction

- I.e. the LP filter can be used for analysis and synthesis
  - analysis
    - input = signal
    - output 1 = prediction coefficients
    - output 2 = 'residual' error
  - synthesis
    - input 1 = prediction coefficients
    - input 2 = 'residual' error
    - output = reconstituted signal
- This makes LP particularly suitable for processing speech signals

The University Of Sheffield.

13

Application to Speech

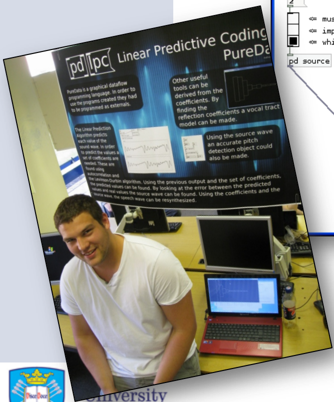
- In speech analysis/resynthesis …
  - the LP filter models the spectral shaping caused by the vocal tract
  - the error sequence models the source excitation
- Note the following assumptions …
  - the analysis model assumes impulse/white-noise excitation
    (*hence the LP filter also captures spectral shaping caused by the glottal source*)
  - the synthesis model ignores any zeros
    (*such as the nasal side branch and the path back down the glottis*)

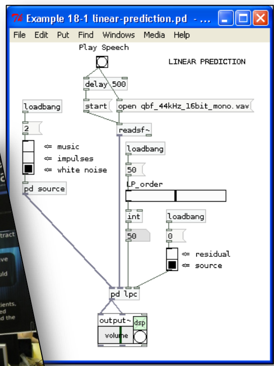
The University Of Sheffield.

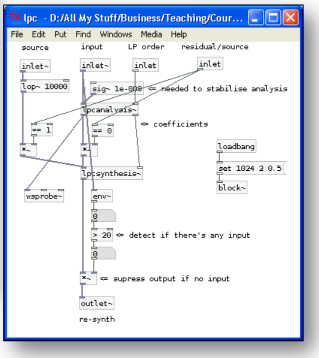
14

Linear Prediction of Speech

pd~







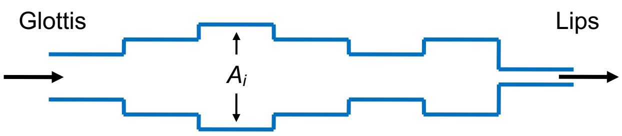
Hassell, A. (2011). ***PD LPC: Linear Predictive Coding in PureData.*** **3rd-Year Project Report, University of Sheffield, Sheffield.**

The University Of Sheffield.

15

Vocal Tract Area Estimation

- Durban's algorithm (*used in the autocorrelation method*) generates parameters that are known as '**reflection coefficients**'
- It can be shown that these are analogous to the proportion of energy reflected at the boundaries of a lossless acoustic tube model of the vocal tract using a set of cylindrical section of equal length
- The reflection coefficients can thus be used to estimate the cross-sectional area of the vocal tract

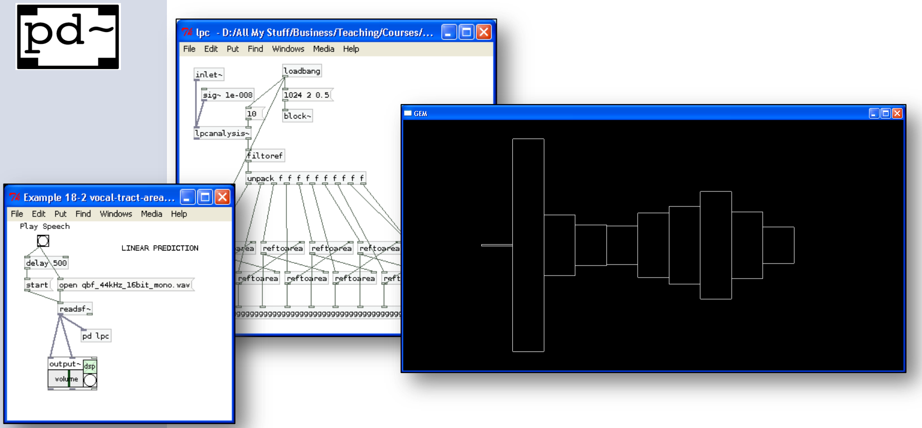


The University Of Sheffield.

16

Vocal Tract Area Estimation

pd~



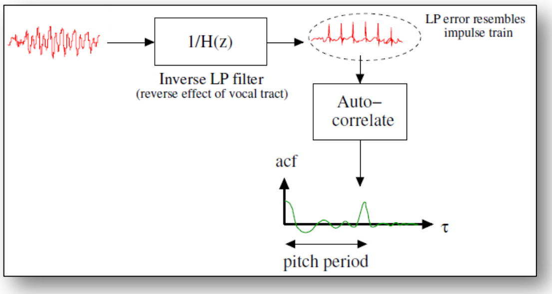
Hassell, A. (2011). *PD LPC: Linear Predictive Coding in PureData.* 3rd-Year Project Report, University of Sheffield, Sheffield.

Applications of Linear Prediction

- pitch extraction
- formant analysis
- speech synthesis
- automatic speech recognition
- linear predictive coding (*LPC*)

LP-based Pitch Extraction

- Obtain LP error signal by passing the speech through the *inverse* filter
- Find the pitch frequency by *autocorrelation* on the error signal



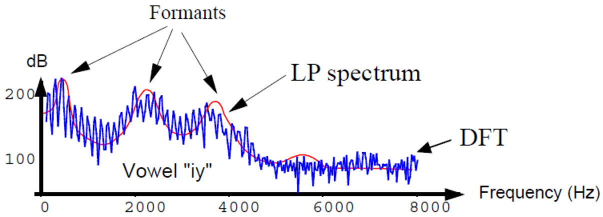
- Use continuity constraints for pitch *tracking*

The University Of Sheffield.

19

LP-based Spectrum Estimation

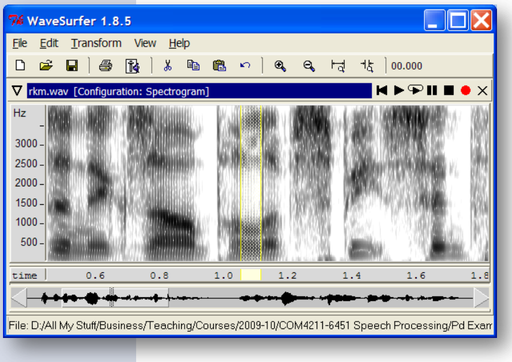
- One way to estimate the speech spectrum is to analyse the frequency response of the LP filter
- Since the excitation information is in the error signal, the LP spectrum is a smoothed all-pole approximation to the speech spectrum
- The LP spectrum is thus useful for calculating spectral features (*such as formant frequencies*)

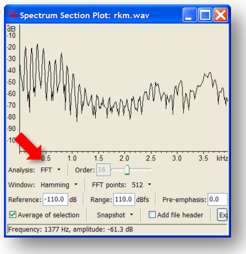


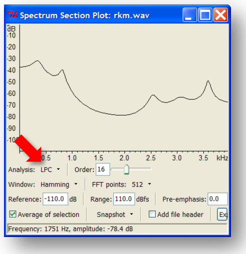
The University Of Sheffield.

20

LP-based Spectrum Estimation







The University Of Sheffield.

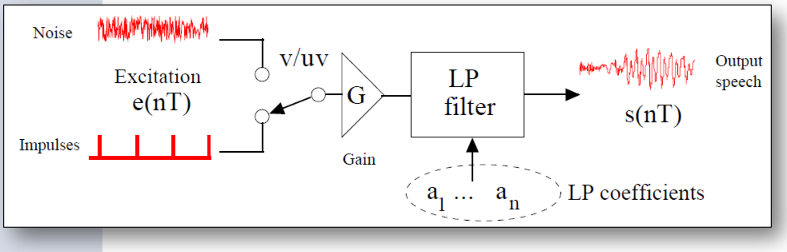
LP-based Formant Extraction

- A pole-pair from LP analysis can model a formant
- There are two main approaches to finding formant candidates …
  - root-solving from the LP polynomial
  - peak-picking from the LP spectrum
- Root-solving finds the frequencies of the pole-pairs with the smallest bandwidths (*i.e. nearest the unit circle*)
- For high-accuracy formant *tracking* it is necessary to consider alternative candidates

The University Of Sheffield.

LP-based Speech Synthesis

An implementation of the *source-filter model*

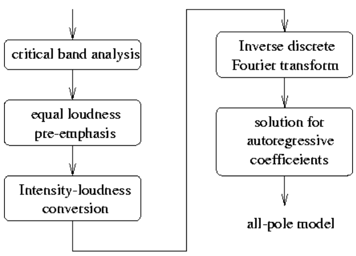


$$s(n) = G.e(n) - a_1 s(n-1) - \ldots - a_p s(n-p)$$

The University Of Sheffield.

LP-based Speech Recognition

- The LP parameters are not guaranteed to be stable from frame-to-frame
- 'Line spectral pairs' (*LSPs*) are a more stable representation derived from the prediction coefficients.
- Some automatic speech recognition systems use 'perceptual linear prediction' (*PLP*) for front-end analysis

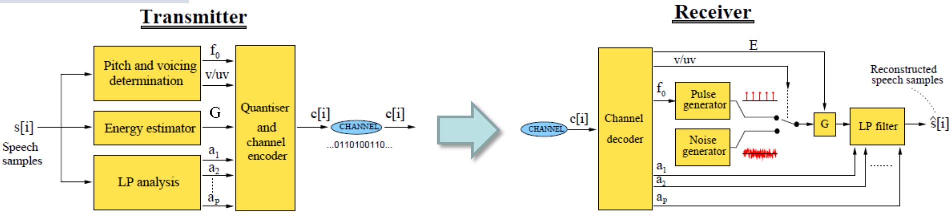


Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. *Journal of the Acoustical Society of America*, 87(4), 1738-1752.

The University Of Sheffield.

Linear Predictive Coding (*LPC*)

- LP *analysis* followed by LP *synthesis*
- Need to quantise and code the excitation and LP filter parameters



- Used in NATO 2.4 kbps '**LPC-10e vocoder**'

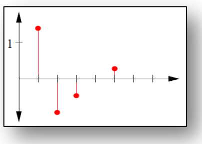
The University Of Sheffield.

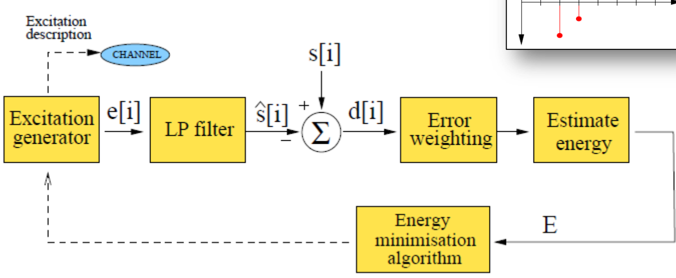
25

Multipulse LPC

Bishnu Atal

- The excitation is set up such that the speech reconstruction error is minimised using an analysis-by-synthesis loop
- The algorithm finds the best pulse position/amplitude, then finds the best position/amplitude of a second, third, etc.
- Pulses can be negative as well as positive





The University Of Sheffield.

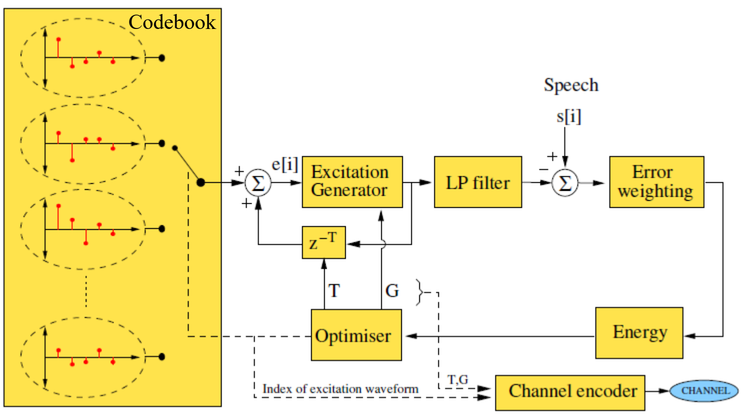
26

Mixed-Excitation LPC (MELP)

- Two-state (*V-UV*) excitation sounds “buzzy”
- Errors in pitch estimation and/or voicing decisions reduce intelligibility
- ‘Mixed-excitation’ models avoid hard V-UV decisions
- MELP has been shown to offer excellent intelligibility at 2400, 1200 and <u>600</u> bps
- The US DoD ‘MELPe’ was adopted in 2002 as the NATO standard vocoder for secure speech communications (*STANAG-4591*)

The University Of Sheffield.

Code-Excited LPC (*CELP*)



- Produces good quality speech over 4.8 kbps
- The basis of most modern speech coding systems

The University Of Sheffield.

GSM Speech Codec

- The digital mobile cellular radio network which is extensively used throughout Europe and many other parts of the world is termed the '**Global System for Mobile communications**' (*GSM*)
- The GSM full rate speech codec operates at 13 kbps
- Basic operating principle …
  - 'CELP'
  - 8 kHz sample rate
  - 20 msec frames
  - 8 short-term predictor coefficients per frame
  - each frame is then further split into four 5 msec sub-frames
  - for each sub-frame the encoder finds a delay and a gain for the codec's long-term predictor
  - the residual signal after both short and long term filtering is quantised for each sub-frame

The University Of Sheffield.

29

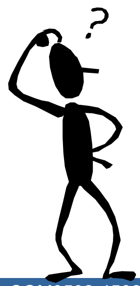
This lecture has covered …

- Linear prediction
- LP analysis of speech signals
- Calculating LP coefficients
- Vocal tract area estimation
- LP-based applications
- Linear predictive coding (*LPC*)

The University Of Sheffield.

30

Any Questions ?



The University Of Sheffield.

31

Next time …

Cepstral Analysis

The University Of Sheffield.

32