ETIN80 — Algorithms in Signal Processors

Projects

Tekn.Dr. Mikael Swartling

Lund Institute of Technology
Department of Electrical and Information Technology

January 22, 2014

Projects

- Speech recognition.
- Speech synthesis.
- Speech separation.
- Adaptive line enhancer.
- Echo cancellation.
- Beat detection.
- Beamforming.
- Or your own project...

Linear Prediction

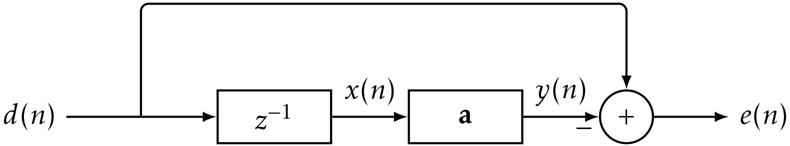
- Given a finite sequence of samples, can we predict the values of future samples?

$$\hat{x}(n+1) = \sum_{k=0}^{K-1} x(n-k)h(k)$$

- Correlated components can be predicted.
  - generally: periodic components
  - fundamentally: tonal components
- Stochastic components cannot be predicted.

Linear Prediction

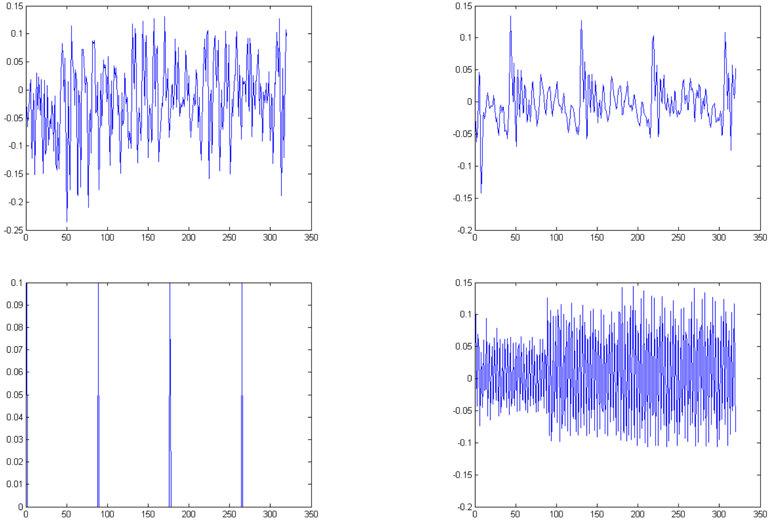
- The 1-step forward linear prediction filter.
- Standard Wiener problem:
  - adaptive solution
  - analytical solution
- The filter describes deterministic properties of the signal.



Linear Prediction

- A linear prediction filter describes formants.
- A formant is an *acoustic resonance*.
- The human vocal tract resonates and generates formants.
  - The vocal cords vibrate at a pitch frequency.
  - Formants generates overtones at different shapes.
  - Fricatives generates shaped noise sounds.
  - A sequence of formats over time describes a “word”.
- The linear prediction error filter cancels the formats.
- The inverse filter oscillates at the formants.

Speech

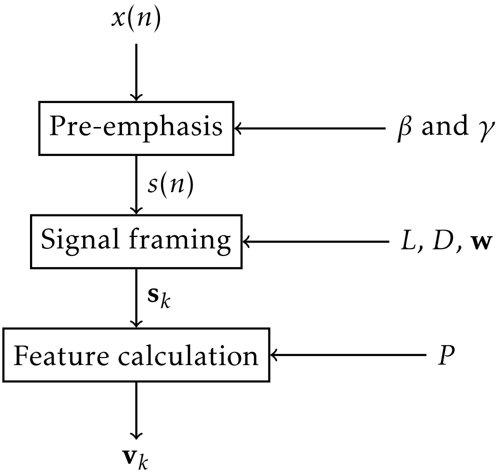

0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
0 50 100 150 200 250 300 350
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
0 50 100 150 200 250 300 350
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 50 100 150 200 250 300 350
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
0 50 100 150 200 250 300 350

Speech

- Speech is stationary for roughly 20 ms.
- Process a speech signal in blocks of 20 ms.
- The filter $\mathbf{a}$ is normalized and pitch-independent.
    - Not dependent on voice amplitude.
    - Not dependent on voice pitch.
    - Different voices have similar filter.

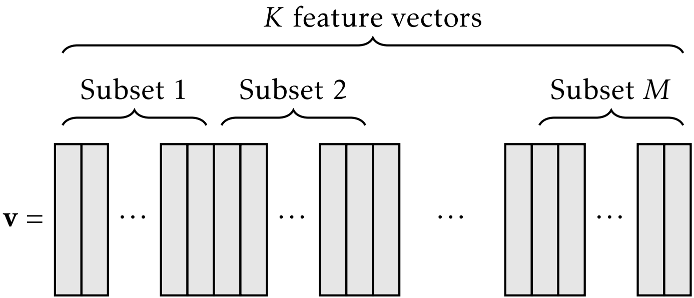
Speech Recognition: Feature Extraction

- Gather a sequence of feature vectors from the speech.



Speech Recognition: Feature Database

- Create an entry from the sequence for future matching.



- The database $\mathbf{d}$ is the set of averages of the M subsets.

Speech Recognition: Matching

$s_0 = 0$

$e_0 = 0$

for each feature vector $\mathbf{v}_k$ in the recorded signal

$d_{\text{curr}} = |\mathbf{v}_k - \mathbf{d}_s|$

$d_{\text{next}} = |\mathbf{v}_k - \mathbf{d}_{s+1}|$

if $d_{\text{curr}} < d_{\text{next}}$

$s_k = s_{k-1}$

else

$s_k = s_{k-1} + 1$

end

$e_k = e_{k-1} + \min(d_{\text{curr}}, d_{\text{next}})$

end

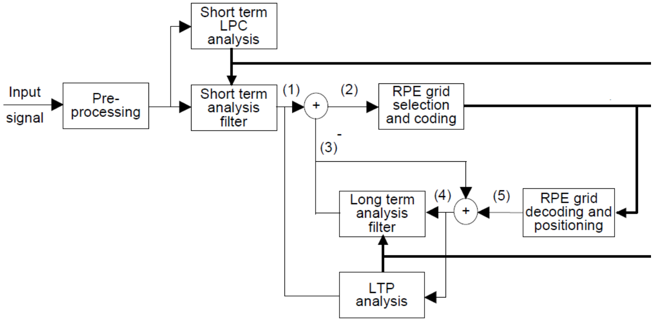
Speech Recognition

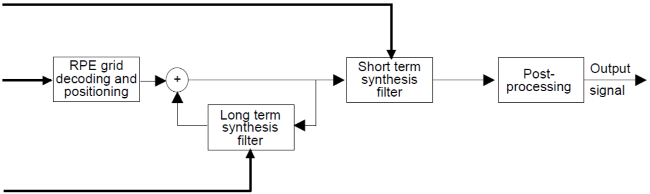
- For single-word verification, detect a match if the error e_k is below a threshold.
- For multi-word lookup, one state machine for every database entry and choose the one with lowest error.

Speech Synthesis

- For every block of a speech signal:
  - Calculate the prediction filter.
  - Calculate the pitch from the error signal.
  - Alter the pitch for funny effects.
  - Construct an excitation signal with the pitch frequency.
  - Filter the excitation signal with the inverse prediction filter.
- A primitive "Vocoder".

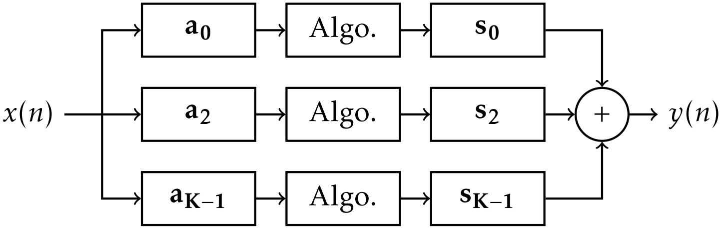
GSM Full-rate Speech Transcoding





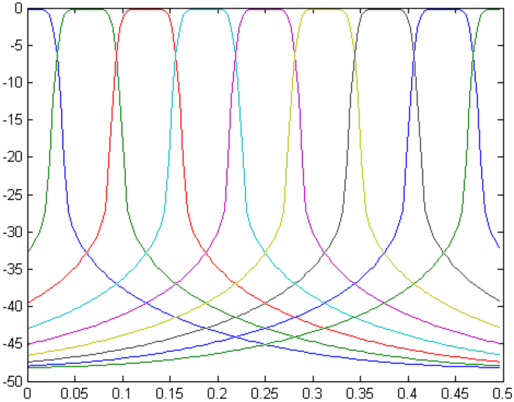
Subband Filtering

- Not all algorithms are suitable for full-band processing.
  - narrow-band algorithm
  - sparse sources and activity detection
- A filterbank is a set of parallel band-pass filters.
  - analysis filter to bandwidth limit the signal
  - subband processing at lower bandwidth
  - synthesis filter for full-band reconstruction



Subband Filtering

- Passbands of 16 subband filters.



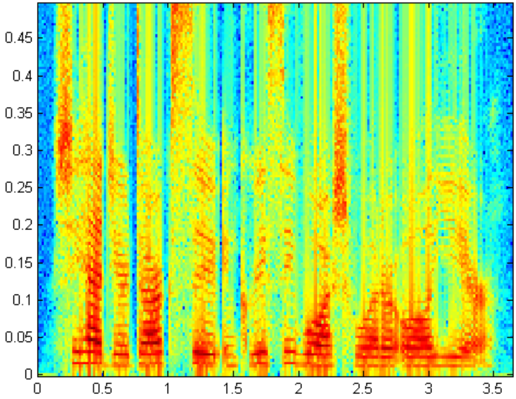
Speech Separation

- Speech excite narrow frequency bands at short times.
- Different speech sources rarely overlap.
- Separate different sources with selective masking.

$$Y_n(\omega,\tau) = \begin{cases} X(\omega,\tau) & \text{If source } n \text{ in } (\omega,\tau). \\ 0 & \text{Otherwise.} \end{cases}$$

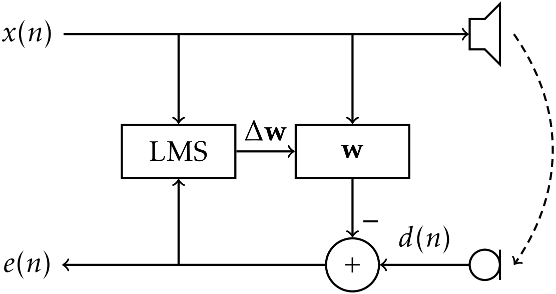
Speech Separation

- Time and frequency sparsity of speech.



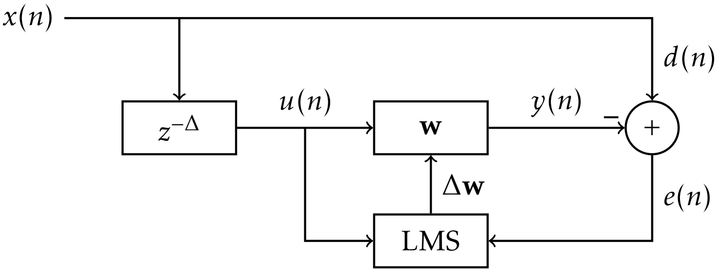
Echo Cancellation

- Cancel the feedback from a speaker into a microphone.
- Adaptive filter for continuous tracking.
- Subband filtering.



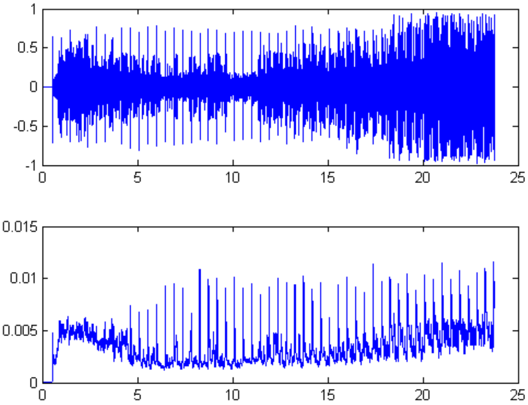
Adaptive Line Enhancer

- Delays a signal by Δ samples and attempts to predict it.
- Cancels uncorrelated components, such as noise.
- Cancels correlated components, such as feedback howling.



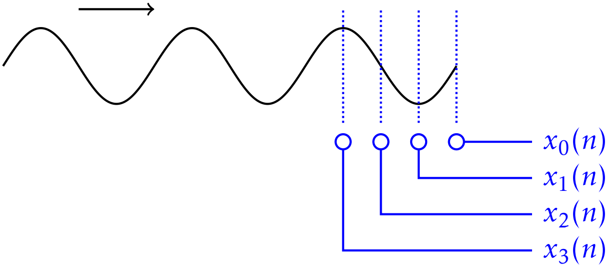
Beat Detection

- The beats in a music dictates its rhythm.
- Mixing of music relies of beat matching.



Beamforming

- Temporal correlation matrix correlates samples in time.
- Spatial correlation matrix correlates samples in space.
- Sampling in time and space are equivalent.

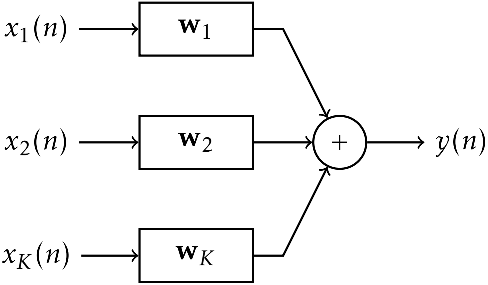


$$\mathbf{x}(n) = \begin{bmatrix} x_0(n) & x_0(n-1) & x_0(n-2) & x_0(n-3) \end{bmatrix} \quad \Rightarrow \mathbf{R_{xx}} = \mathbf{x}^T\mathbf{x}$$

$$\mathbf{x}(n) = \begin{bmatrix} x_0(n) & x_1(n) & x_2(n) & x_3(n) \end{bmatrix} \quad \Rightarrow \mathbf{R_{xx}} = \mathbf{x}^T\mathbf{x}$$

Beamforming

- Filtering in both temporal and spatial domain.
- Listening to or cancel a spatially located source.
- Standard methods such as $\mathbf{w}_{\text{opt}} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{d\mathbf{x}}$ still applies.



Choose your own project

Or, choose your own project...