Signal Analysis Using Autoregressive Models of Amplitude Modulation

Sriram Ganapathy

Advisory - Hynek Hermansky
Johns Hopkins University
11-18-2011
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - *product of smooth modulation with a fine carrier.*
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.

\[ x(t) = m(t) \times \cos\{\omega_0 t + \varphi(t)\} \]
Desired Properties of AM

- **Linearity**
  \[ \alpha x(t) \quad \Rightarrow \quad \alpha m(t) \]

- **Continuity**
  \[ x(t) + \delta x(t) \quad \Rightarrow \quad m(t) + \delta m(t) \]

- **Harmonicity**
  \[ \cos(\omega_0 t) \quad \Rightarrow \quad 1 \]
Desired Properties of AM

- Uniquely satisfied by the analytic signal

\[
\begin{align*}
H & \quad x_a(t) \\
\mathcal{H} & \quad x(t) \\
\omega(t) & \quad \frac{d}{dt} \\
\omega_0 + \phi(t) & \quad \angle \\
m(t) & \quad || \\
\end{align*}
\]

\( \mathcal{H} \) - Hilbert transform, \( x_a(t) \) - analytic signal,
\( |x_a(t)|^2 \) - Hilbert envelope
Desired Properties of AM

- However, the Hilbert transform filter is infinitely long and can cause artifacts for finite length signals.

\[
\mathcal{H}(x(t)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t-\tau)}{t-\tau} \, d\tau
\]

- Need for modeling the Hilbert envelope without explicit computation of the Hilbert transform.
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
AR Model of Hilbert Envelopes

Signal $x[n]$ with zero mean in time and frequency domain for $n = 0 \ldots N-1$

Discrete-time analytic spectrum

$$X_a[k] = \begin{cases} 2X[k] & \text{for } k < N/2 \\ 0 & \text{for } k \geq N/2 \end{cases}$$
AR Model of Hilbert Envelopes

Signal $x[n]$ with zero mean in time and frequency domain for $n = 0 \ldots N-1$

Discrete-time analytic spectrum

$$X_a[k] = \begin{cases} 
2X[k] & \text{for } k < N/2 \\
0 & \text{for } k \geq N/2 
\end{cases}$$
Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for $n < N$, $q[n] = x[M - n], M = 2N - 1$

Spectrum

$Q[k] = 2Re\{X[k]\}$
AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$. $q[n] = x[n]$ for $n < N$, $q[n] = x[M - n]$, $M = 2N - 1$

Discrete-time analytic spectrum

$$Q[k] = 2Re\{X[k]\}$$

$$Q_a[k] = \begin{cases} 2Q[k], & k < N \\ 0, & k \geq N \end{cases}$$
AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.  
$q[n] = x[n]$ for $n < N$,  
$q[n] = x[M - n], M = 2N - 1$

Discrete-time analytic spec.  
\[
Q[k] = 2\text{Re}\{X[k]\}
\]
\[
2Q[k], \quad k < N
\]
\[
Q_a[k] = 0, \quad k \geq N
\]

N-point DCT  
\[
y[k] = 4\text{Re}\{X[k]\}, \quad k < N
\]
AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$. 
$q[n] = x[n]$ for $n < N$, $q[n] = x[M - n]$, $M = 2N - 1$

Discrete-time analytic spec.

$Q[k] = 2Re\{X[k]\}$

$Q_a[k] = \begin{cases} 
2Q[k], & k < N \\
0, & k \geq N 
\end{cases}$

DCT zero-padded with N-zeros

$\hat{y}[k] = \begin{cases} 
4Re\{X[k]\}, & k < N \\
0, & k \geq N 
\end{cases}$
AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for $n < N$, $q[n] = x[M - n]$, $M = 2N - 1$

Discrete-time analytic spec.

\[
Q[k] = 2\text{Re}\{X[k]\}
\]

\[
Q_a[k] = \begin{cases} 
2Q[k], & k < N \\
0, & k \geq N 
\end{cases}
\]

DCT zero-padded with $N$-zeros

\[
\overline{y[k]} = \begin{cases} 
4\text{Re}\{X[k]\}, & k < N \\
0, & k \geq N 
\end{cases}
\]

\[
Q_a[k] = \mathcal{F}\{q_a[n]\} = \overline{y[k]}
\]
We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = \overline{y[k]} \]
We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = \overline{y[k]} \]
AR Model of Hilbert Envelopes

We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = \overline{y[k]} \]
AR Model of Hilbert Envelopes

We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = \overline{y[k]} \]

Even-sym. analytic spectrum.

Zero-padded DCT sequence.
AR Model of Hilbert Envelopes

We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = y[k] \]

\[ \mathcal{F}\{|q_a[n]|^2\} = r_y[\tau] \]

Spectrum of Hilbert env. for even-sym. signal

Auto-correlation of DCT sequence
AR Model of Hilbert Envelopes

We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = y[k] \]

\[ \mathcal{F}\{|q_a[n]|^2\} = r_y[\tau] \]

Hilb. env. of even-symm. signal

\[ \mathcal{F} \]

Auto-corr. of DCT
AR Model of Hilbert Envelopes

We have shown -

\[ Q_a[k] = \mathcal{F}\{q_a[n]\} = y[k] \]

\[ \mathcal{F}\{|q_a[n]|^2\} = r_y[\tau] \]
LP in Time and Frequency
LP in Time and Frequency

- **Time** → **LP** → **Power Spec.**
- **DCT** → **LP** → **Hilb. Env.**
FDLP

Linear prediction on the cosine transform of the signal

Speech
FDLP

Linear prediction on the cosine transform of the signal
FDLP

Linear prediction on the cosine transform of the signal
FDLP

Linear prediction on the cosine transform of the signal

Speech

FDLP Env.

Hilb. Env.
FDLP for Speech Representation
FDLP for Speech Representation

DCT
FDLP for Speech Representation
FDLP for Speech Representation

DCT → LP

Diagram showing the process from DCT to LP.
FDLP for Speech Representation
FDLP for Speech Representation

FDLP Spectrogram

Freq.

Time
FDLP for Speech Representation

FDLP Spectrogram

Conventional Approaches
FDLP versus Mel Spectrogram

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Resolution of FDLP Analysis

Sig.

FDLP Env.
Resolution of FDLP Analysis

\[ \text{Res.} = \frac{1}{\text{Critical Width}} \]
Resolution of FDLP Analysis

Sig. Length = 125ms
p = 16
Resolution of FDLP Analysis

Sig. Length = 125ms
p = 16

Location of First Peak (ms)
Properties of FDLP Analysis

• Summarizing the gross temporal variation with a few parameters
  – Model order of FDLP controls the degree of smoothness.
  – AR model captures perceptually important high energy regions of the signal.

• Suppressing reverberation artifacts
  – Reverberation is a long-term convolutive distortion.
    • Analysis in long-term windows and narrow sub-bands.
Properties of FDLP Analysis

• Summarizing the gross temporal variation with a few parameters
  – Model order of FDLP controls the degree of smoothness.
  – AR model captures perceptually important high energy regions of the signal.

• Suppressing reverberation artifacts
  – Reverberation is a long-term convolutive distortion.
    • Analysis in long-term windows and narrow sub-bands.
Reverberation

When speech is corrupted with convolutive distortion like room reverberation

\[
\text{Clean Speech} \ast \text{Room Response} = \text{Revb. Speech}
\]
Reverberation

When speech is corrupted with convolutive distortion like room reverberation

\[
\text{Clean Speech} \ast \text{Room Response} = \text{Revb. Speech}
\]

In the long-term DFT domain, this translates

\[
\text{Clean DFT} \times \text{Response DFT} = \text{Revb. DFT}
\]
Reverberation

When speech is corrupted with convolutive distortion like room reverberation

\[ r[n] = x[n] \ast h[n] \]

In the DFT domain, this translates to a multiplication

\[ R[k] = X[k] \times H[k] \]

In the \( m^{th} \) sub-band,

\[ R_m[k] = X_m[k] \times H_m[k] \]
Reverberation

$H[k]$
Reverberation

$H[k]$
Reverberation

$H[k]$
Reverberation

\[ H[k] \]

\[ H_m \]
Reverberation

When speech is corrupted with convolutive distortion like room reverberation

\[ r[n] = x[n] * h[n] \]

In the DFT domain, this translates to a multiplication

\[ R[k] = X[k] \times H[k] \]

In the \( m^{th} \) sub-band,

\[ R_m[k] = X_m[k] \times H_m[k] \]

In narrow bands, \( H_m[k] \) is approx. constant,

\[ R_m[k] \approx X_m[k] \times H_m \]
Gain Normalization in FDLP

- FDLP envelope of $m^{th}$ band using all-pole parameters \( \{a_1, ..., a_p\} \) is given by

\[
\hat{E}_m[n] = \frac{G}{|1 - \sum_{k=1}^{p} a_k e^{-j2\pi kn/N}|^2}
\]

- When the sub-band signal is multiplied by $H_m$, the gain $G$ is modified.

- **Normalization** to convolutive distortions is achieved by reconstructing the FDLP envelope with $G = 1$. 
Gain Normalization in FDLP

(a) Without gain norm.

(b) With gain norm.

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Outline of Applications

- Sub-band Decomposition
- FDLP
- Gain Norm.
- Quant.
- AM
- FM
- Short-term Features for Speaker & Speech Recog.
- Modulation Features for Phoneme Recog.
- Wide-band Speech & Audio Coding

Input Signal

Short-term Features

Sub-band Decomposition

FDLP

Gain Norm.

Quant.

Short-term Features for Speaker & Speech Recog.

Modulation Features for Phoneme Recog.

Wide-band Speech & Audio Coding
Short-term Features

Input → DCT → Sub-band Window → FDLP → Gain Norm. → Energy Int. → Log + DCT → Features
Envelopes in each band are integrated along time (25 ms with a shift of 10 ms).
Integration in frequency axis to convert to mel scale.
- Sub-band energies are converted to cepstral coefficients by applying log and DCT along frequency axis.
- Delta and acceleration coefficients are appended to obtain 39 dim. feat similar to conventional MFCC feat.
Speech Recognition

- TIDIGITS Database (8 kHz)
  - Clean training data, test data can be clean or naturally reverberated.
- HMM-GMM system
  - Whole-word HMM models trained on clean speech.
  - Performance in terms of word error rate (WER).
- Features
  - PLP features with cepstral mean subtraction (CMS).
  - Long-term log spectral sub. (LTLSS) [Avendano],[Gelbart]
Speech Recognition

Speaker Verification

- NIST 2008 Speaker recognition evaluation (SRE)
  - Has telephone speech and far-field speech.

- GMM-UBM system
  - Trained on a large set of development speakers.
  - Adapted on the enrollment data from the target speaker.
  - Nuisance attribute projection (NAP) on supervectors.
  - Detection cost function (DCF) = 0.99 $P_{fa} + 0.1 P_{miss}$

- Features with warping [Pelecanos, 2001].
  - Mel Frequency Cepstral Coefficients (MFCCs)
  - FDLP short-term (FDLP-S) features.
Speaker Verification

Outline of Applications

- Sub-band Decomposition
- FDLP
- Gain Norm.
- Quant.
- Short-term Features for Speaker & Speech Recog.
- Modulation Features for Phoneme Recog.
- Wide-band Speech & Audio Coding
Modulation Features

- Sub-band Decomposition
- FDLP
- Gain Norm.
- Quant.

Output:
- AM
- FM

- Short-term Features for Speaker & Speech Recog.
- Modulation Features for Phoneme Recog.
- Wide-band Speech & Audio Coding
Modulation Feature Extraction

![Diagram showing the process of modulation feature extraction involving DCT, critical-band window, FDLP, and DCT for static and dynamic components, all followed by 200ms and leading to sub-band features.]
- Static compression is a logarithm – reduce the huge dynamic range in the sub-band envelope.
Dynamic compression is implemented by dynamic compression loops consisting of dividers and low pass filters [Kollmeier, 1999].
- Compressed sub-band envelopes are DCT transformed to obtain modulation frequency components
- 14 static and dynamic modulation spectra (0-35 Hz) with 17 sub-bands, gets a feature of 476 dim.
Phoneme Recognition

- TIMIT Database (8 kHz)
  - Clean training data, test data can be clean, additive noise, reverberated or telephone channel.
- Multi-layer perceptron (MLP) based system
  - MLPs estimate phoneme posteriors
  - Hidden Markov model (HMM) – MLP hybrid model.
  - Performance in phoneme error rate (PER).
- Features
  - Perceptual linear prediction (PLP) - 9 frame context.
  - FDLP modulation (FDLP-M) features – 476 dim.
Phoneme Recognition

Outline of Applications

- Sub-band Decomposition
- FDLP
- Gain Norm.
- Quant.
- Short-term Features for Speaker & Speech Recog.
- Modulation Features for Phoneme Recog.
- Wide-band Speech & Audio Coding

Input Signal

Diagram flow:
- Sub-band Decomposition
- FDLP
- Gain Norm.
- Quant.
- Output:
  - Short-term Features
  - Modulation Features
  - Wide-band Features
Audio Coding

Sub-band Decomposition

FDLP

Gain Norm.

Quant.

AM

FM

Input Signal

Short-term Features for Speaker & Speech Recog.

Modulation Features for Phoneme Recog.

Wide-band Speech & Audio Coding
Subjective Evaluations

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary
Summary

- Employing **AR modeling** for estimating amplitude modulations.

- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.

- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.
Summary

- Employing AR modeling for estimating amplitude modulations.

- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.

- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.
Summary

- Employing AR modeling for estimating amplitude modulations.

- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.

- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.
Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.

- Investigating the resolution properties of FDLP.

- Gain normalization of FDLP Envelopes
Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.

- Investigating the resolution properties of FDLP.

- Gain normalization of FDLP Envelopes
Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.
- Investigating the resolution properties of FDLP.
- Gain normalization of FDLP Envelopes
Our Contributions

- **Short-term feature** extraction using FDLP – Improvements in reverb speech recog.

- **Modulation feature extraction** – Phoneme recognition in noisy speech.

- **Speech and audio codec development** using AM-FM signals from FDLP.
Our Contributions

- **Short-term feature** extraction using FDLP – Improvements in reverb speech recog.

- **Modulation feature** extraction – Phoneme recognition in noisy speech.

- Speech and audio codec development using AM-FM signals from FDLP.
Our Contributions

- **Short-term feature** extraction using FDLP – Improvements in reverb speech recog.

- **Modulation feature** extraction – Phoneme recognition in noisy speech.

- **Speech and audio codec development** using AM-FM signals from FDLP.
Publications

Journals


Patents

Temporal Masking in Audio Coding Based on Spectral Dynamics in Frequency Sub-bands

"Spectral Noise Shaping in Audio Coding Based on Spectral Dynamics in Frequency Sub-bands
Selected Conferences


Acknowledgements

- **Lab Buddies** – Samuel Thomas, Sivaram Garimella, Padmanbhan Rajan, Harish Mallidi, Vijay Peddinti, Thomas Janu, Aren Jansen.

- **Idiap personnel** – Petr Motlicek, Joel Pinto, Mathew Doss.

- **IBM personnel** – Jason Pelecanos, Mohamed Omar

- **Others** – Xinhui Zhou, Daniel Romero, Marios Athineos, David Gelbart, Harinath Garudadri.
Thank You
Noise Compensation in FDLP

- When speech is corrupted with additive noise,

\[ y[n] = x[n] + s[n] \]

- The noise component is additive in the non-parametric Hilbert envelope domain (assuming the signal and noise are uncorrelated).
- Voice activity detector (VAD) provides information about the non-speech regions which are used for estimating the temporal envelope of the noise.
- Noise subtraction tries to subtract the estimate the noise envelope from the noisy speech envelope.
Noise Compensation in FDLP

Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.
Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.
Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.

- Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].

- Gain normalization deals with short and long term distortions within a single long-term frame.
Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.
Feature Comparison

MFCC

- clean
- tel.

Frames

FDLP

- clean
- tel.

Frames

- clean
- revb.

Frames

- clean
- revb.
Evidences

- Physiological evidences -
  - Spectro-temporal receptive fields [Shamma et.al. 2001]

- Psycho-physical evidences -
  - Perceptual importance of modulation frequencies [Drullman et al. 1994].
  - Syllable recognition from temporal modulations with minimal spectral cues [Shannon et al., 1995].
Evidences

- Physiological evidences -
  - Spectro-temporal receptive fields [Shamma et al. 2001].

- Psycho-physical evidences -
  - Perceptual importance of modulation frequencies [Drullman et al. 1994].
  - Syllable recognition from temporal modulations with minimal spectral cues [Shannon et al., 1995].
Applications

- Modulation spectra has been used in the past
  - Speech intelligibility [Houtgast et al, 1980].
  - RASTA processing [Hermansky et al, 1994].
  - Speech recognition [Kingsbury et al, 1998].
  - AM-FM decomposition [Kumaresan et al, 1999].
  - Sound texture modeling [Athineos et al, 2003].
  - Sound source separation [King et al, 2010].
Linear Prediction – Time Domain

- Current sample expressed as a linear combination of past samples
Linear Prediction – Time Domain

- Current sample expressed as a linear combination of past samples

\[ x[n] = \sum_{k=1}^{p} a_k x[n-k] + e[n] \quad \forall \ n = 0 \ldots N - 1 \]

- Model parameters are solved by minimizing the residual sum of squares.

\[ E_p = \sum_{n=0}^{N-1} |e[n]|^2 \]
AR model of Power Spectrum

Filter interpretation [Makhoul, 1975]

\[ e[n] = x[n] - \sum_{i=1}^{p} a_i x[n - i] = x[n] * d[n] \]

\[ d = [1 - a_1 - a_2 \ldots - a_p] \]

\[ \mathcal{E}(\omega) = \sum_{n=0}^{N-1} e[n] e^{-j\omega n} = X(\omega)D(\omega) \]

From Parseval’s theorem

\[ E_p = \sum_{n=0}^{N-1} |e[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathcal{E}(\omega)|^2 \, d\omega \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 |D(\omega)|^2 \, d\omega \]
AR model of Power Spectrum

By definition,

\[ |D(\omega)|^2 = |1 - \sum_{i=1}^{p} a_i e^{-ji\omega}|^2 \]

Let,

\[ P_x(\omega) = |X(\omega)|^2, \quad H(\omega) = \frac{1}{D(\omega)} \]

Thus, parameters \( \{a_i\} \) are solved by minimizing

\[ E_p = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 |D(\omega)|^2 \ d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_x(\omega)}{|H(\omega)|^2} \ d\omega \]
Solution of the linear prediction yields an all-pole model of the power spectrum

\[
\widehat{P}_x[\omega] = Ep |H(\omega)|^2 = \frac{G}{|1 - \sum_{i=1}^{p} a_i e^{-j\omega}|^2}
\]

Numerator \( G \) denotes the gain of AR model (equal to minimum residual sum of squares).
AR model of power spectrum
Hilbert Envelope - Definition

- **Analytic signal** is the sum of the signal and its quadrature component.

\[ x_a[n] = x[n] + j\mathcal{H}(x[n]) \]

where \( \mathcal{H} \) denotes the Hilbert transform.

- **Hilbert envelope** is the squared magnitude of the analytic signal.
LP in Time and Frequency

- **Magnitude (dB)**
  - **Power Spectrum**
  - **LP Spectrum**

- **Frequency (Hz)**
  - Hertz range from 0 to 4000 Hz

- **Time (ms)**
  - Milliseconds range from 0 to 35 ms

- **Magnitude (dB)**
  - Values range from 0 to 110 dB
  - Red line for LP Spectrum
  - Blue dots for Power Spectrum

- **Hilbert Envelope**
  - Blue dots

- **FDLP Envelope**
  - Red line
AM-FM Decomposition

(a) Signal
(b) Hilb. Env.
(c) FDLP Env.
(d) AM comp.
(e) FM comp.
Modulation Feature Extraction

Input \(\rightarrow\) DCT \(\rightarrow\) Critical-band Window \(\rightarrow\) FDLP \(\rightarrow\) Static \(\rightarrow\) DCT \(\rightarrow\) Sub-band Feat.
Dynamic \(\rightarrow\) DCT

200ms
Modulation Features

(a)

- Signal
- Hilb. Env.
- FDLP Env.
- Log comp.
- Dyn. comp.

Conventional signal analysis – starts with the estimation of **short-term spectrum** (10-40 ms).
Introduction

- Conventional signal analysis – starts with the estimation of short-term spectrum (10-40 ms).

- Spectrum is sampled at a preset rate before further modeling/processing stages.

- Contextual information is typically processed with time-series models such as HMM.
Introduction

- Conventional signal analysis – starts with the estimation of **short-term spectrum** (10-40 ms).

- Spectrum is **sampled at a preset rate** before further modeling/processing stages.

- **Contextual information** is typically processed with time-series models such as HMM.